Molecular Genetic Analysis of Indene Bioconversion in *Rhodococcus* strain I24

by

Sheri L. Treadway

B. S. *summa cum laude* Microbiology
Arizona State University, 1993

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGY
AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 1999

Signature of Author: .....
Department of Biology
April 22, 1999

Certified by: ..
Anthony J. Sinskey
Professor of Microbiology
Thesis Supervisor

Accepted by: .....
Alan D.Grossman
Professor of Biology
Committee for Graduate Students

Science

Table of Contents

Molecular Genetic Analysis of Indene Bioconversion in *Rhodococcus* strain I24

by

Sheri L. Treadway

Submitted to the Department of Biology
On April 22, 1999 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in Biology

Abstract

A bioconversion is a stereospecific chemical transformation using whole cells such as bacteria as the catalyst for the reaction(s). Bioconversions are commonly used in industrial production of compounds such as antibiotics, vitamins, and precursors to therapeutics. Bacteria from the genus *Rhodococcus* have been shown to be hardy, solvent tolerant, and an excellent source of stereospecific enzymes. Therefore, Rhodococci are ideal candidates for bioconversions. *Rhodococcus* strain I24 was identified by its ability to utilize aromatic hydrocarbons such as naphthalene and toluene as sole carbon sources. This strain can perform a bioconversion of the aromatic hydrocarbon indene to two enantiomers of 1,2-indandiol that can be used in the synthesis of the HIV-1 protease inhibitor Crixivan™. The bioconversion of indene by *Rhodococcus* strain I24 is a complex network modeled as three parallel pathways that produces a variety of indene derived compounds. Metabolic engineering, the use of genetic engineering combined with chemical engineering, can be used to analyze this bioconversion process. In this work the genetic engineering aspects of metabolic engineering are described. The indene bioconversion pathways were analyzed genetically and biochemically. We cloned and sequenced genes encoding one branch of the indene bioconversion network. This branch, encoded by the genes *nidABC*, includes a naphthalene-like dioxygenase system (*nidAB*) and a diol dehydrogenase (*nidC*). The NidAB dioxygenase can catalyze the conversion of indene to *cis*-(1R,2S)-indandiol, 1-indenol, and 1-indandone. The NidC dehydrogenase can oxidize *cis*-(1R,2S)-indandiol to keto-hydroxy-indan. Biochemical analysis of the dehydrogenase activity demonstrated that the NidC dehydrogenase is stereospecific for *cis*-(1R,2S)-indandiol. Since we were interested in dissecting the indene bioconversion network, the properties of a *nidC* null mutant were investigated. Results from these analyses suggest the presence of a second diol dehydrogenase that is naphthalene-inducible in *Rhodococcus* strain I24 capable of utilizing *cis*-(1R,2S)-indandiol as a substrate. The impact of these data on our model for the indene bioconversion network is discussed. The genetic analysis of this system serves as a framework for the genetic engineering component of metabolic engineering.

Thesis Supervisor: Anthony J. Sinskey
Title: Professor of Microbiology

For my dearest family and friends

Acknowledgements

Many thanks go to my thesis advisor, Dr. Anthony J. Sinskey, who was instrumental in guiding me on the critical path. His encouragement, wisdom, and humor have been invaluable. Special thanks go to Dr. Alan Grossman who has been extremely helpful in the development of my scientific career. I would also like to thank all the members of my committee including Drs. Alan Grossman, Robert Sauer, Graham Walker, and Stephen Drew for their interest in my work and their guidance.

The members of the Sinskey Laboratory are greatly appreciated for helpful discussions, technical assistance, and most of all for providing a fun atmosphere in which to work (think Phil singing Disney songs). Thanks to Phil Lessard for sharing a bay with me and being an all around great guy, to Xian O'Brien for transforming from student to friend, to Laura Willis for allowing me to ramble on about this and that, and to Stephane Guillouet for his driving skills and great sense of humor. Thanks also goes to the plethora of Urops who provided Black Magic Cake (Molly Hills), entertaining stories (Adrianne Chapman), assistance for acquiring things from a great height (Alison Higgins), and entertainment and potato chips (Elaine Lankenau and Irene Kim)

I would also like to thank those from a time past including Kevin "Is he Beavis or Butthead?" Mills, Neal "The Good Critic" Silverman, Jun "Tough to say" Horiuchi, and Greg "I'm out of time!" Marcus. And of course Ruth and Josie for their garlic pasta and special care.

All of this would not have been possible without the support and encouragement of my family. Thanks to my parents who have always encouraged me to do my best. And especially to Jim, who has a budding scientific career as "Filter Boy", with whom I intend to spend the rest of my life.

Chapter One: Introduction

Human Immunodeficiency Virus, HIV-1

The human immunodeficiency virus (HIV-1), previously called human T-cell leukemia/lymphoma virus III (HTLV-III), was first discovered in the early 1980's (Barre-Sinoussi et al. 1983; Gallo et al. 1983; Sarngadharan et al. 1984). This retrovirus is responsible for the symptoms associated with Acquired Immunodeficiency Syndrome, better known as AIDS. HIV-1 infection has become prevalent worldwide and AIDS has become a pandemic. Over the last fifteen years the HIV-1 retrovirus has been extensively studied at the molecular level to better understand the virus, its components and the mechanism of pathogenesis. Consequently, drugs capable of combating the virus have been developed based on our understanding of the viral components and the viral life cycle.

Genome Structure and Life Cycle of HIV-1

Retroviruses are distinguished from other viruses by the presence of an RNA genome that is reverse-transcribed by viral enzymes and integrated into the host DNA. There are two types of retroviruses, simple and complex; HIV-1 is categorized as a complex retrovirus. Complex retroviruses are distinguished from the simple viruses because they have regulatory factors such as transcription activators and other factors that enhance/assist at different stages of the viral life cycle. The genome of the HIV-1 virus was sequenced in 1984 by two separate groups and found to be approximately 10 kb in length (Hahn et al. 1984; Luciw et al. 1984). Since then numerous laboratories have analyzed the genome and its constituent parts in great detail (for detailed reviews see (Frankel and Young 1998; Katz and Skalka 1994; Klotman and Wong-Staal 1991)). A model for the organization of the viral genome and the relation of these parts to the assembled viral particle is depicted in Figure 1-1. The HIV-1 genome has three major open reading frames called *gag, pol,* and *env* that encode the structural components. These open reading frames produce the three major polyproteins produced by the virus, the Gag polyprotein, the Gag-Pol polyprotein, and the Env polyprotein. These polyproteins are cleaved during the life cycle of the virus to produce the major components of the virus. The Gag polyprotein contains the matrix protein (MA), the capsid protein (CA), the nucleocapsid (NC), and a fourth protein called p6. The Gag-Pol polyprotein produces the protease (PR), the reverse transcriptase (RT), and the integrase (IN). The Env polyprotein produces two components of a membrane protein, the surface protein (SU/gp120) and the transmembrane protein (TM/gp41). In addition to the structural components and viral enzymes produced from the polyproteins, the genome also encodes a number of regulatory factors including Tat, Rev, Nef, Vif, Vpr, and Vpu.

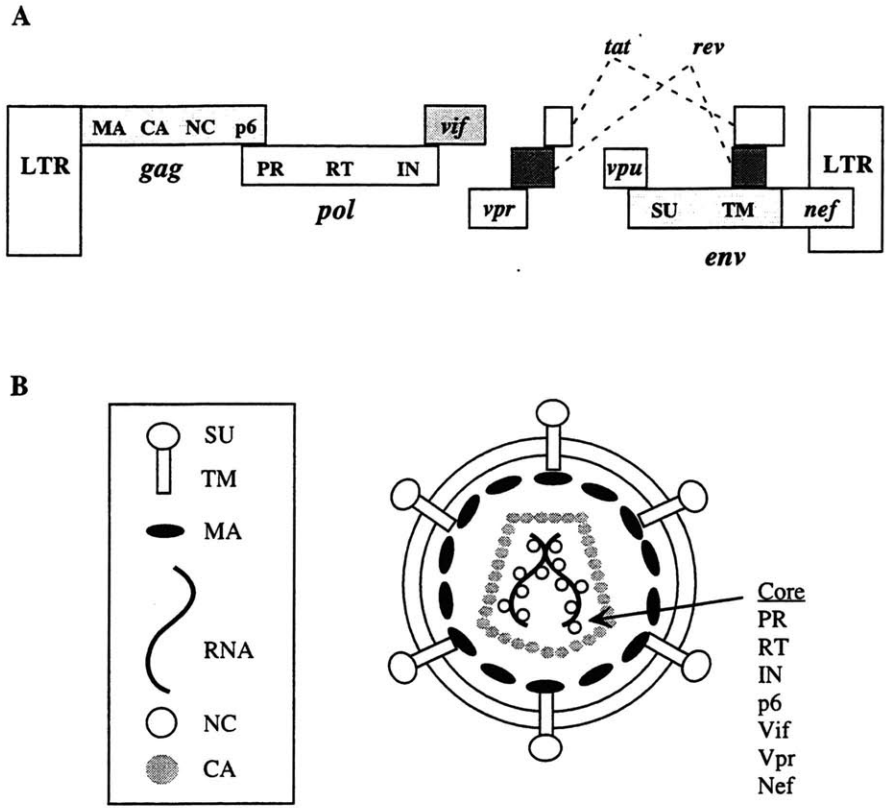


Figure 1-1: HIV-1 Genome Structure and Virus Particle Structure
A) The HIV-1 genome structure based on sequence information and molecular analysis of the components. Three major open reading frames (*gag, pol,* and *env*) encode the major structural components and enzymes. The regulatory and auxiliary factors are encoded separately and many are the result of splicing events of the mRNA. The symbols of the proteins in the major open reading frames are matrix (MA), capsid (CA), nucleocapsid (NC), p6, protease (PR), reverse transcriptase (RT), integrase (IN), surface protein (SU) and transmembrane protein (TM) B) Structure of the HIV-1 virion particle. The membrane is derived from the host cell during budding of the immature virus particle.

The life cycle of the retrovirus is depicted in Figure 1-2 and can be broken down into three sections: virus entry, replication, and virus packaging and maturation. The life cycle begins with virus entry into the host cell. In the case of HIV-1 the host cell is either a T cell or a macrophage. The surface protein (SU) binds to receptors (CD4) and co-receptors (chemokine receptors) on the host cell surface. After binding, the transmembrane protein (TM) undergoes a conformational change that promotes cell fusion. The outer membrane of the virus particle is removed exposing the virion core and the RNA is converted to partially double stranded DNA by the reverse transcriptase (RT). The core is then transported across the nuclear membrane with assistance from the integrase enzyme (IN) and the auxiliary protein Vpr. After entry into the nucleus the integrase catalyzes the integration of the viral derived DNA into the host chromosome.

The next stage of the life cycle is replication of the virus which begins with the production of viral transcripts using the host cell's RNA polymerase II from the promoter of the viral genome found in the 5'-long terminal repeat (LTR). Transcription is greatly enhanced by the transcription activator Tat that binds the Tat Activating Region (TAR) found in the promoter. The RNA can remain whole producing mRNA that will give rise to the Gag polyprotein, the Gag-Pol polyprotein or the RNA for the new viral particles. In addition, the RNA can also be spliced to produce mRNA for the Env polyprotein and mRNAs for the various auxiliary proteins. This stage is completed when the RNA is transported out of the nucleus and into the cytoplasm. Transport is regulated by the auxiliary protein Rev.

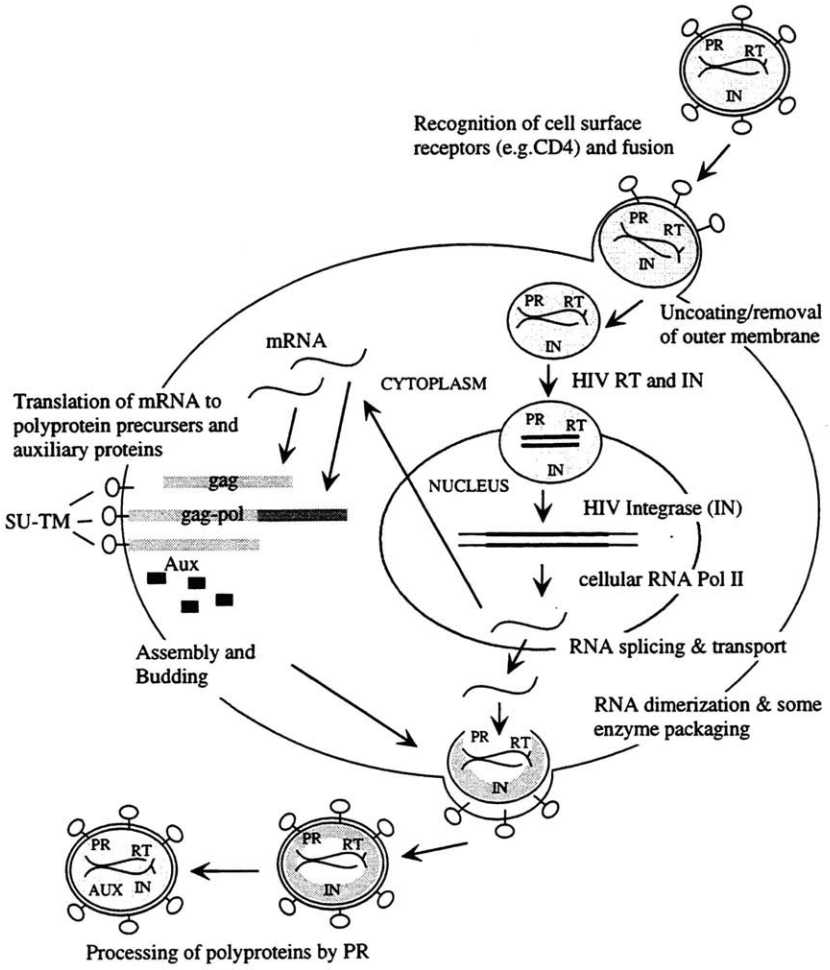


Figure 1-2: HIV-1 Life Cycle
The life cycle of the virus is depicted starting from virus entry into the cell (based on diagrams from (Frankel and Young 1998; Katz and Skalka 1994; Klotman and Wong-Staal 1991)). Auxiliary proteins (AUX) include the transcription activator Tat, the regulator Rev, and other factors including Vpu, Vpr, Vif, and Nef.

The final stage, virus packaging and maturation, begins with the translation of the mRNAs into the various proteins or polyproteins. The Env polyprotein is transferred to the endoplasmic reticulum and

undergoes endoproteolytic cleavage to produce the two constituent proteins, the surface (SU) and transmembrane (TM) proteins. A disulfide-bond associates the two proteins and the complex associates with CD4 to target it to the cell surface. During provirus assembly the Vpu and Nef proteins promote release of the surface protein complex from CD4. The Gag-Pol polyprotein is produced by a frameshift of the ribosome during translation. The polyproteins Gag and Gag-Pol are targeted for assembly into the provirus by packaging signals and are responsible for capturing the RNA for provirus packaging. During provirus packaging a small quantity of protease, reverse transcriptase, and integrase are packaged into the provirus. After packaging is complete, the provirus buds from the host cell and becomes a mature virion when the polyproteins are cleaved by the protease (PR) producing the viral enzymes and capsid components.

The HIV-1 Protease and Protease Inhibitors

The HIV-1 protease is a key enzyme in the maturation of the virus particle to its infectious state. The protease is responsible for the cleavage of the polyprotein precursors of the major capsid proteins and enzymes. It is an aspartyl protease that is similar to other aspartyl proteases such as renin and endothiopepsin. The functional enzyme is a dimer with each monomer contributing the catalytic residues (Asp-Thr/Ser-Gly) to the symmetric catalytic site. The protease uses a general acid-base mechanism and cleaves between Tyr-Pro residues or Phe-Pro residues. No known consensus cleavage site has been determined for retroviral proteases (Katz and Skalka 1994).

The protease is a prime target for drug therapy to combat HIV infection. Inhibition of the protease should inhibit the life cycle of the virus by preventing virion maturation, thereby reducing the viral load in a patient. It has been shown that chemical or mutational inactivation of the HIV-1 protease leads to the production of immature, non-infectious virus particles (Kohl et al. 1988; McQuade et al. 1990; Seelmeier et al. 1988). A number of protease inhibitors have been developed in the last ten years. These therapeutic agents were produced in rational drug design efforts by a variety of companies using the structure of the protease as a basis for their designs. Numerous crystal structures of the protease have been determined of the protease alone and the protease complexed with other molecules (for review see (Wlodawer and Vondrasek 1998)). Currently there are four protease inhibitors that have been approved for use by the Food and Drug Administration (FDA). These structurally similar inhibitors include Saquinavir (Hoffman-La Roche), Ritonavir (Abbott), Nelfinavir (Agouron Pharmaceuticals), and Indinavir (Merck) (Vacca and Condra 1997).

Indinavir - A HIV-1 Protease Inhibitor

Indinavir, also known as Crixivan™, is one of four commercially available HIV-1 protease inhibitors. In clinical trails this protease inhibitor is very effective in combination therapy with two reverse transcriptase inhibitors, zidovudine and lamivudine (Hammer et al. 1997; Rutschmann and Hirschel 1997). The structure of the inhibitor is shown in Figure 1-3 (Buckland et al. 1999). This inhibitor is a transition state

analog that is reversible; it has also been called a peptidomimetic inhibitor. That is, the compound resembles/mimics a certain state, the transition state, of the polyprotein during the cleavage reaction (Wlodawer and Vondrasek 1998). Indinavir has five chiral centers; therefore, there are thirty-two possible conformations of this compound. Only one of these confirmations is the active form of the therapeutic agent. The compound is currently manufactured in a chemical synthesis by Merck and Company, Inc. which requires numerous separate chemical steps. In each step involving a chiral center significant loss of active product can occur (Buckland et al. 1999; Reider 1997).

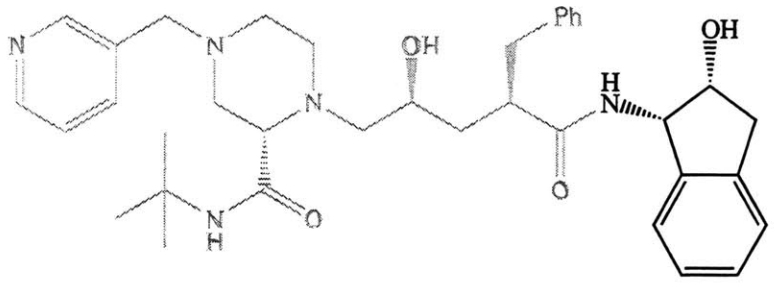


Figure 1-3: Structure of the HIV-1 Protease Inhibitor Crixivan™
The structure of the protease inhibitor indinavir sulfate is depicted above (Buckland et al. 1999). An intermediate, *cis*-amino-indanol, in the synthesis of the compound is highlighted in black.

Chiral Synthesis of Therapeutics

Many therapeutics are chiral in nature and often only one enantiomer is active. Crixivan is one such compound. Therefore, stereospecific synthesis is required to produce many therapeutic agents. Bacterial enzymes can contribute to the chiral synthesis of therapeutic compounds or precursor molecules (synthons). Enzymes from bacteria are often highly efficient and capable of carrying out stereospecific reactions in a single step, where as a chemical synthesis may require multiple steps due to the protecting and unprotecting of reactive centers. Bacterial enzymes have been used in two ways to produce chiral synthons. One way is to isolate the bacterial enzyme and use it in an *in vitro* reaction. The second method requires the use of whole cells in what is called a bioconversion. A bioconversion, also called a biotransformation, is the process of using organisms such as bacteria to conduct chemical transformations. Often the bioconversion substrate can not support growth of the organism.

There are numerous examples of drugs, vitamins, and precursor compounds that are produced using a bioconversion at some stage in production. For example, vitamin C is produced by combining chemical steps with a bioconversion in a genetically engineered strain of *Acetobacter* (Glazer and Nikaido 1995). A precursor of the anti-inflammatory compound Naproxen is produced in a *Bacillus subtilis* bioconversion (Patel 1997). Other compounds such as prednisone (a type of cortisone) (Primrose 1991), precursors to

Taxol (Patel 1997; Zaks and Dodds 1997) and vitamin B_T (Zimmermann et al. 1997) also utilize bioconversions in their synthesis.

It may be possible to produce a precursor to the HIV-1 protease inhibitor Crixivan™ using a bioconversion process. The aromatic hydrocarbon indene could be used as the initial substrate for this bioconversion. The hypothesized reactions necessary would be an oxygenation reaction and an amino-transferase reaction (Figure 1-4). Two oxygenated versions of indene, *cis*-(1S,2R)-indandiol and *trans*-(1R,2R)-indandiol, can be used as precursors for Crixivan™ (Buckland et al. 1999). The addition of the amino group to these precursors could be conducted in an *in vitro* reaction, or possibly engineered into a bioconverting strain. The first reaction (reaction A) that could produce these compounds could be conducted by an oxygenase similar to those found in bacteria capable of degrading aromatic compounds. Aromatic hydrocarbon metabolism/degradation has been extensively studied in bacteria.

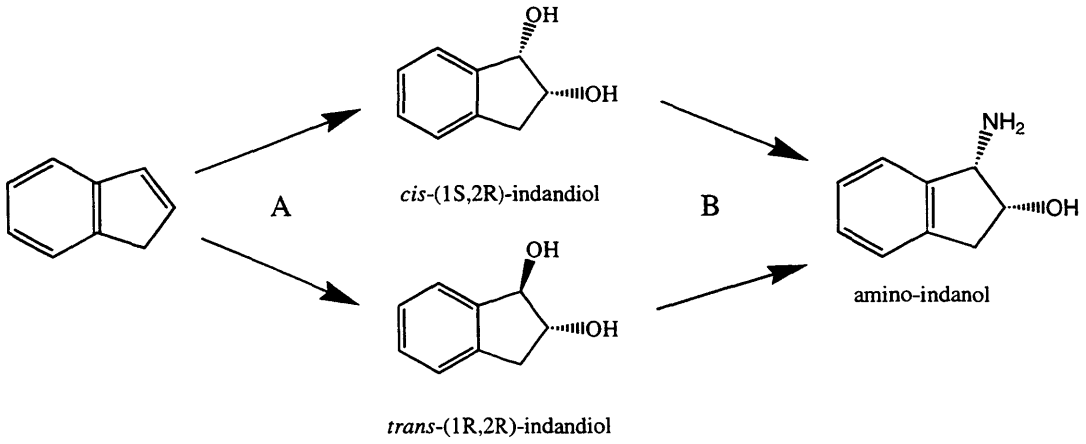


Figure 1-4: Potential bioconversion pathway for the synthesis of Crixivan™
This pathway shows two routes to the intermediate amino-indanol. Enzyme A would be a type of oxygenase. Enzyme B would be an amino-transferase. It is possible that a microbe could carry out this bioconversion or part of this bioconversion naturally. Alternatively, a microbe could be engineered to carry out this bioconversion.

Bacterial Aromatic Hydrocarbon Degradation

Aromatic hydrocarbon compound degradation/catabolism pathways in bacteria allow the organism to utilize novel compounds as carbon sources rendering the bacteria capable of occupying numerous unique environments. The first aromatic degrader, a strain of *Bacillus haxacarbovorum,* was identified in 1908 by its ability to utilize toluene and xylene as carbon sources (for a historical review see (Gibson and Subramanian 1984)). Over the years a wide range of Gram-negative and Gram-positive bacteria have been identified that can metabolize many different aromatic hydrocarbons. Among these organisms are *Pseudomonas*, *Rhodococcus*, *Sphingomonas,* and *Alcaligenes,* and *Moraxella* (Dagley 1971; Gibson and

Subramanian 1984; Mueller et al. 1997; Warhurst and Fewson 1994). Strains have been found that can degrade toluene, xylene, naphthalene, benzene, isopropyl benzene, chlorinated aromatic hydrocarbons such as PCB's, and dibenzofuran among others (Bhat and Vaidyanathan 1995; Butler and Mason 1997; Grund et al. 1995).

Aromatic compound degradation/catabolism can be carried out under aerobic conditions or anaerobic conditions by very different pathways. Only aerobic pathways will be discussed here. Aerobic degradation is comprised of two pathways, the upper pathway and the lower pathway (Kiyohara et al. 1994; Sanseverino et al. 1993; Williams and Sayers 1994). Two example pathways containing an upper and a lower pathway are shown in Figures 1-5 and 1-6. Figure 1-5 shows one route to metabolize toluene and Figure 1-6 shows a pathway for naphthalene metabolism. The upper pathway consists of the first steps of degradation through ring cleavage to produce catechol, a common intermediate of these pathways. An oxygenase system (either a monooxygenase or dioxygenase) initiates the upper pathway, followed by an oxidation step conducted by a dehydrogenase. The final portion of the upper pathway is a second oxygenation by a different oxygenase that also catalyzes cleavage of the aromatic ring. If the compound is a poly-aromatic compound, such as naphthalene, there will be multiple oxygenation and ring cleavage steps in order to reach catechol (see Figure 1-6).

The lower pathway follows from the upper pathway by continuing to breakdown the ring-cleavage product so that it may be ultimately used in primary carbon metabolism. The lower pathway can be subcategorized into the *meta* pathway and the *ortho* pathway (also called the β-ketoadipate pathway). The type of lower pathway is dependent on the position of the ring-cleavage that occurs relative to the recently added hydroxyl groups (Williams and Sayers 1994). As new strains of bacteria are identified and studied, numerous aromatic hydrocarbon degradation/metabolism pathways have been elucidated. The extensive study of these pathways allows for some generalizations about the components of these pathways. Specifically, components of the upper pathway including oxygenases and dehydrogenases will be discussed both biochemically and genetically. These enzymes could play a role in a putative bioconversion of indene to produce a chiral synthon.

Figure 1-5: Toluene Catabolism/degradation Pathway
One route for toluene catabolism by *Pseudomonas putida* is shown (Zylstra et al. 1988). The upper and lower pathways are demarcated. The upper pathway contains a dioxygenase system with multiple electron transport proteins (see below).

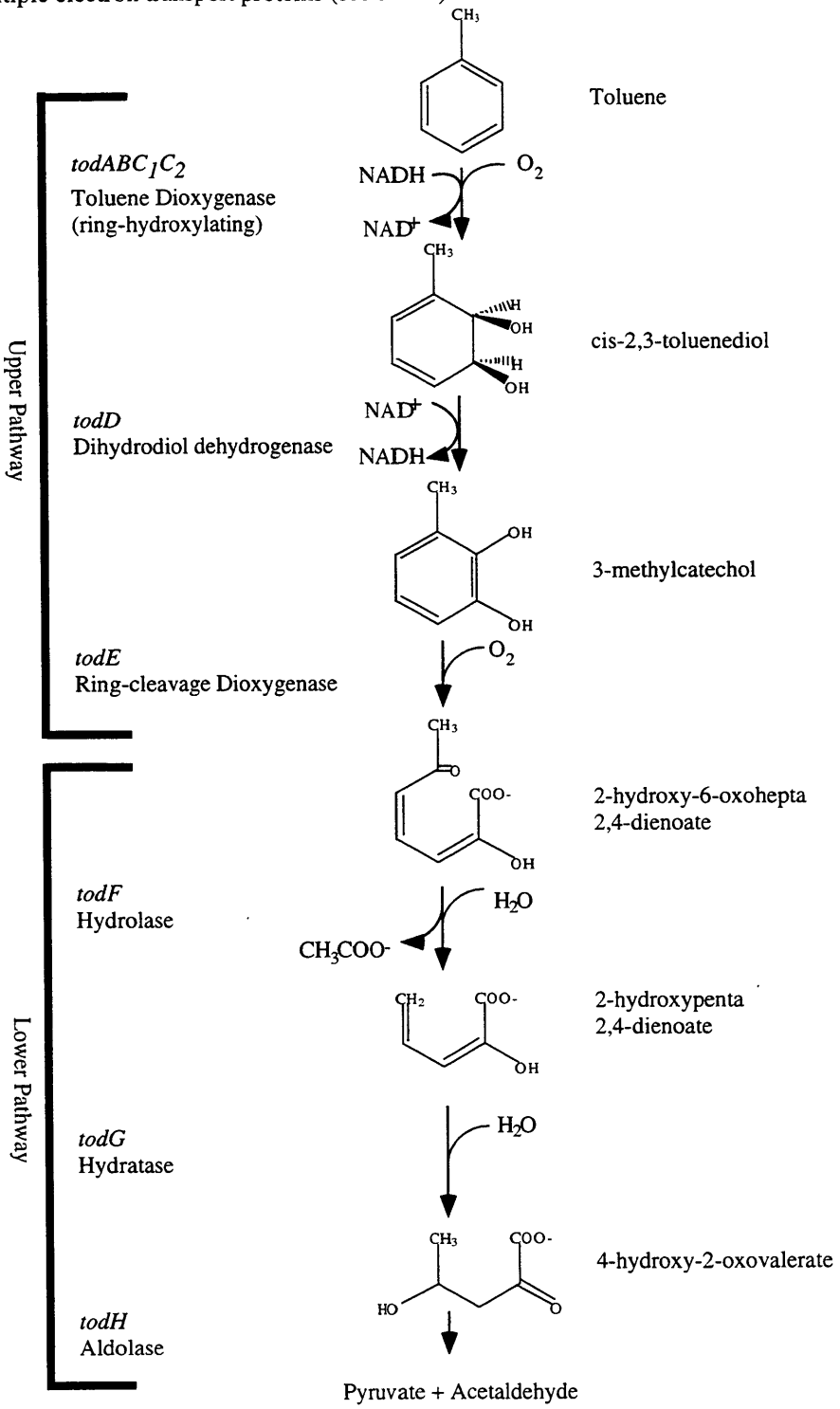
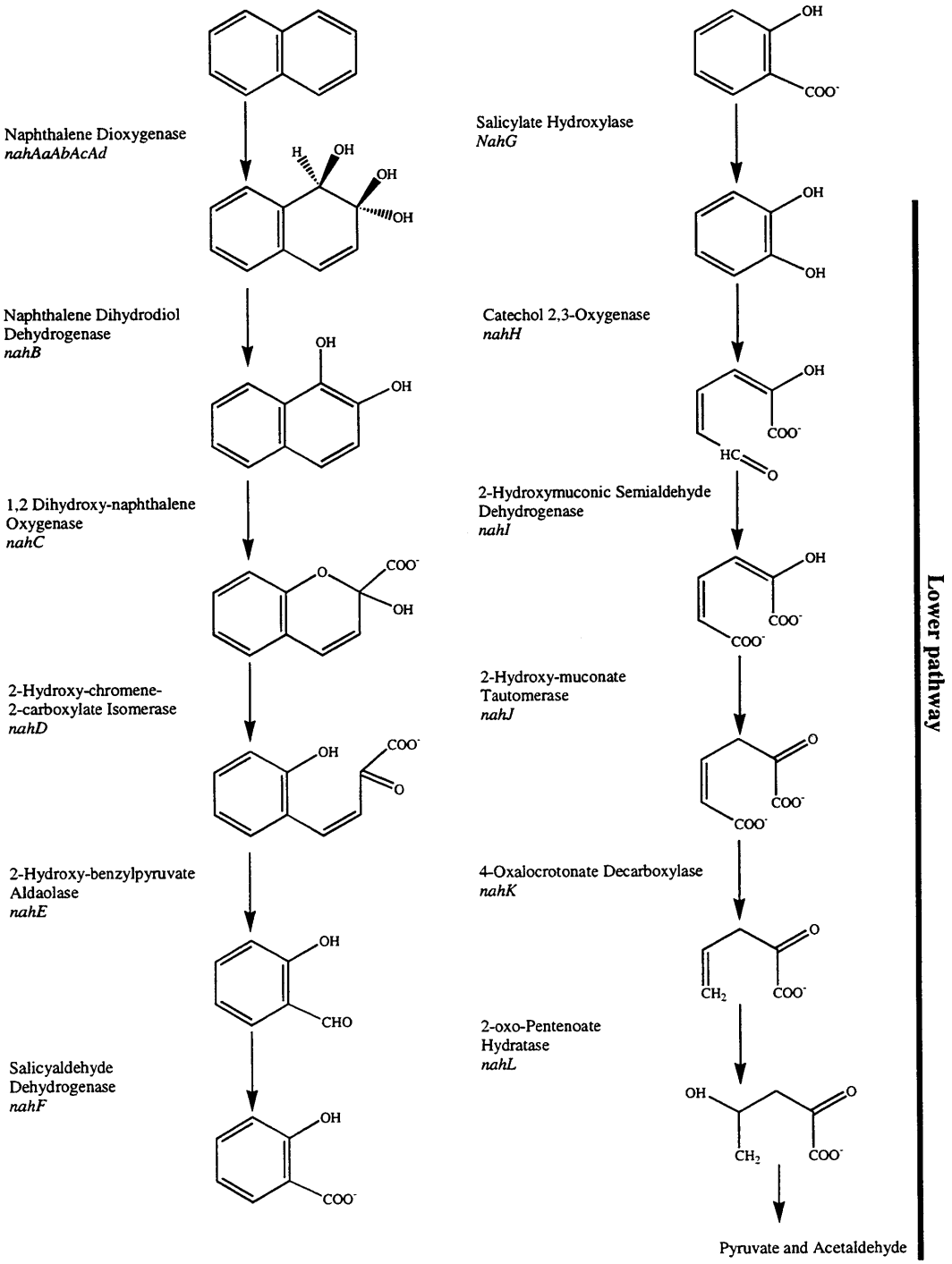


Figure 1-6: Naphthalene Catabolism/degradation Pathway

Naphthalene catabolism *Pseudomonas putida* is shown (Zylstra 1994). This pathway has two parts, an upper and a lower pathway. The upper pathway is extended and contains a dioxygenase system with multiple electron transport proteins (see below). The lower pathway is marked with a line.



Oxygenases

There are two main types of oxygenases, monooxygenases and dioxygenases. In general, monooxygenases incorporate one atom of oxygen from molecular oxygen and dioxygenases incorporate both atoms of oxygen (Hayaishi 1962). Monooxygenases are also known as mixed function oxidases or hydrolases (Hayaishi 1962; Mason 1988). The distinction between monooxygenases and dioxygenases, however, is not always clear. There are examples of monooxygenases that oxygenate repeatedly producing a product that appears to have resulted from a dioxygenation reaction (Lei and Tu 1996; Oldfield et al. 1997). Furthermore, some dioxygenases have monooxygenation activities depending on the substrate (Brand et al. 1992; Gibson et al. 1995; Lehning et al. 1997; Spain et al. 1989; Wackett et al. 1988). Consequently, the classification of oxygenases as monooxygenases or dioxygenases is only suggestive of their function, but not their full potential as enzymes. In addition not only can oxygenases add molecular oxygen, but they can dehalogenate simultaneously as in PCB degradation (chlorinated aromatics) or degradation of other halogenated compounds (Fetzner 1998; Khan and Walia 1989), and remove sulfur or nitrogen (Gray et al. 1996; Spain 1995).

There are many different types of oxygenases based upon the type of cofactors they employ. All oxygenases contain a cofactors such as transition metals, heme, and/or flavin or pteridine that interact with molecular oxygen (Harayama et al. 1992; Hayaishi 1962). For example, many oxygenases use Fe(II) or Fe(III) as cofactors (Harayama et al. 1992). Other oxygenases such as the chlorocatechol 1,2-dioxygenase from *Rhodococcus erythropolis* 1CP contain manganese as a cofactor (Maltseva et al. 1994), and a *Pseudomonas* L-tryptophan 2, 3-dioxygenase contains copper as the metal cofactor (Feigelson 1976). They may also have iron-sulfur centers (2Fe-2S centers) in addition to metal cofactors (Harayama et al. 1992). Those oxygenases that contain flavin as a cofactor are also called flavoproteins. Flavin-containing oxygenases are generally comprised of a single polypeptide and frequently are monooxygenases (Harayama et al. 1992).

As mentioned above, some oxygenases are comprised of a single polypeptide. Other oxygenases require a short electron transport chain for their activity. Both monooxygenases and some dioxygenases fall into this category. The electron transport chain is generally comprised of two redox centers: the first containing a flavin and the second containing an iron-sulfur center. The redox centers are either in one electron transport protein, or two separate proteins. The configuration of the redox centers varies with the system. A schematic for electron transport systems is depicted in Figure 1-7. In general, electron transport proteins acquire their electrons from NAD(P)H. This electron is then passed to the terminal electron acceptor, the oxygenase, simultaneously activating the enzyme. For the rest of this chapter only bacterial dioxygenases will be discussed.

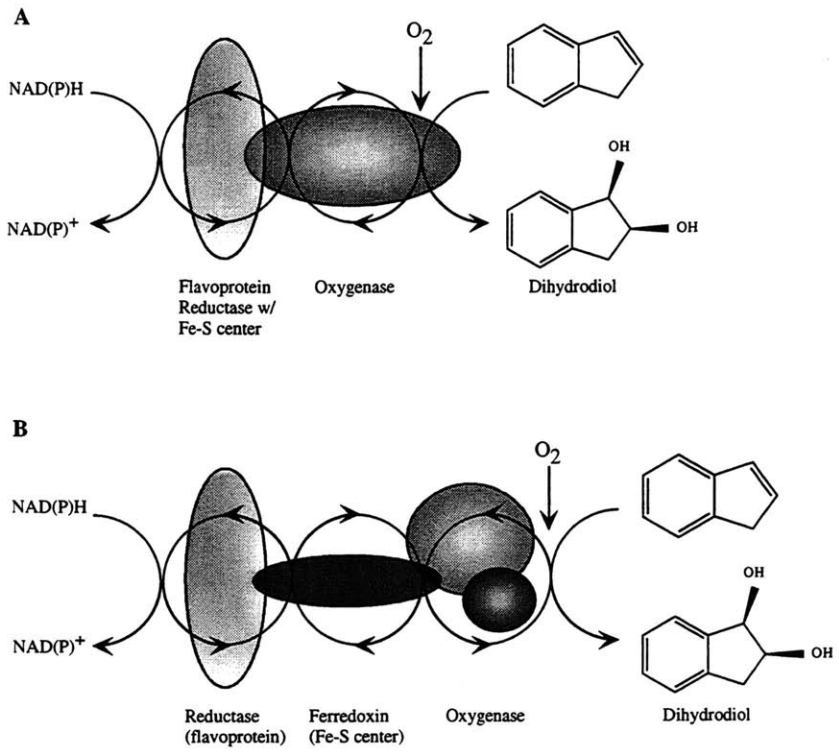


Figure 1-7: Electron Transport System of Oxygenases
Two electron transport systems are depicted that can either be for a monooxygenase or a dioxygenase system. A) Single electron transport protein system, B) Multiple electron transport protein system.

Bacterial Dioxygenases

There are two types of bacterial dioxygenases, ring-hydroxylating dioxygenases and ring-cleavage dioxygenases (Butler and Mason 1997; Harayama et al. 1992; Mason and Cammack 1992). The ring-hydroxylating dioxygenases incorporate two hydroxyl groups into the aromatic ring and require an electron transport chain. The ring-cleavage dioxygenases likewise incorporate two hydroxyl groups into the aromatic substrate, but they also catalyze ring cleavage and no electron transport chain is required. There are two types of ring-cleavage dioxygenases, the intradiol (*ortho*) dioxygenases and the extradiol (*meta*) dioxygenases. Intradiol dioxygenases cleave the aromatic ring between the two hydroxyl groups that were added to the ring. Extradiol dioxygenases are cleaved outside of the hydroxyl groups (Figure 1-8). The combined activities of the ring-hydroxylating dioxygenase and the ring-cleavage dioxygenase are a major portion of the "upper pathway" of the aromatic compound catabolism by bacteria. These enzymes help prepare the substrate for entry into lower pathway and central carbon metabolism. For a review of ring-cleavage dioxygenases see the article by Harayama (Harayama et al. 1992). Only the ring-hydroxylating dioxygenases will be discussed henceforth.

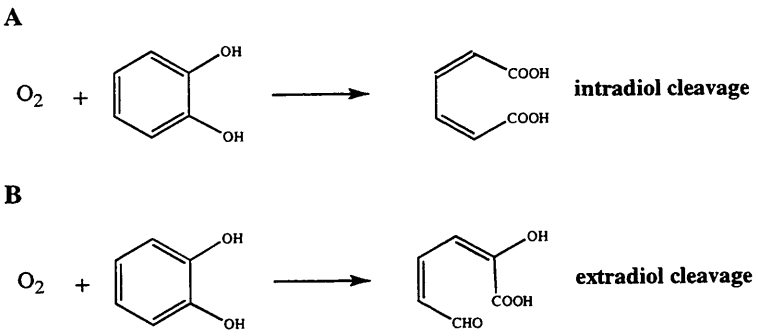


Figure 1-8: Two types of Ring-cleavage Dioxygenase
There are two types of ring-cleavage dioxygenases, intradiol and extradiol. The distinction between the two types is shown diagrammatically here. A) intradiol cleavage, B) extradiol cleavage.

Ring-hydroxylating dioxygenases are tightly associated oligomers of a single subunit (α_n) or two subunits ($\alpha_n\beta_n$) that require an electron transport chain. All of the bacterial enzymes known to date are non-heme enzymes that contain Rieske-type 2Fe-2S centers and one mononuclear iron cofactor that is generally found in the α subunit (Ensley and Gibson 1983; Kauppi et al. 1998; Yamaguchi and Fujisawa 1982). Sometimes the iron-sulfur center is produced with components from both subunits (Mason and Cammack 1992). The mononuclear iron binding site is also the active site of the enzyme.

There are three main classes of ring-hydroxylating dioxygenases (Batie et al. 1991; Butler and Mason 1997). This classification is based on the configuration of the electron transport chain and the characteristics of the electron transport proteins. Figure 1-9 shows a schematic representation of the different classes of dioxygenases with examples. Class I dioxygenase systems have two major components: the flavoprotein reductase that contains both redox centers of the electron transport chain (a flavin and a Fe-S center) and the dioxygenase. There are two subclasses, Class IA and Class IB. Class IA has a flavoprotein reductase that has flavin mononucleotide (FMN) as the cofactor and an iron-sulfur center. Class IB has flavin adenine dinucleotide (FAD) and an iron-sulfur center. Some examples of Class I systems include the phthalate dioxygenase from *Pseudomonas cepacia* (Batie et al. 1987), 4-sulphobenzoate 3,4-dioxygenase from *Comamonas testosteroni* (Locher et al. 1991), and 2-halobenzoate 1,2-dioxygenase from *P. cepacia* (Fetzner et al. 1992). Unlike Class I systems, Class II dioxygenase systems are three component systems comprised of an oxygenase and two electron transport proteins. Each of the electron transport proteins, the reductase component and the ferredoxin component, contain a redox center. In all cases the reductase contains FAD as the cofactor and the ferredoxin contains the iron-sulfur center. There are also two subclasses of Class II, Class IIA and Class IIB, which are differentiated by the type of ferredoxin. Class IIA dioxygenase systems have a ferredoxin with an iron-sulfur center. Pyrazon dioxygenase from *Pseudomonas* is one such system (Sauber et al. 1977). Class IIB systems, on the other hand, have a Rieske-type iron-sulfur center; Benzene dioxygenase (Geary et al. 1990) and biphenyl dioxygenase (Haddock and Gibson 1995) are known to be part of this class of

enzymes. As with Class II systems, Class III dioxygenase systems have three components; the only one known to date is the naphthalene dioxygenase system of *Pseudomonas putida.* This system is comprised of the terminal oxygenase, a reductase flavoprotein that also contains an iron-sulfur center, and a ferredoxin that contains a Rieske-type iron-sulfur center (Haigler and Gibson 1990a; Haigler and Gibson 1990b).

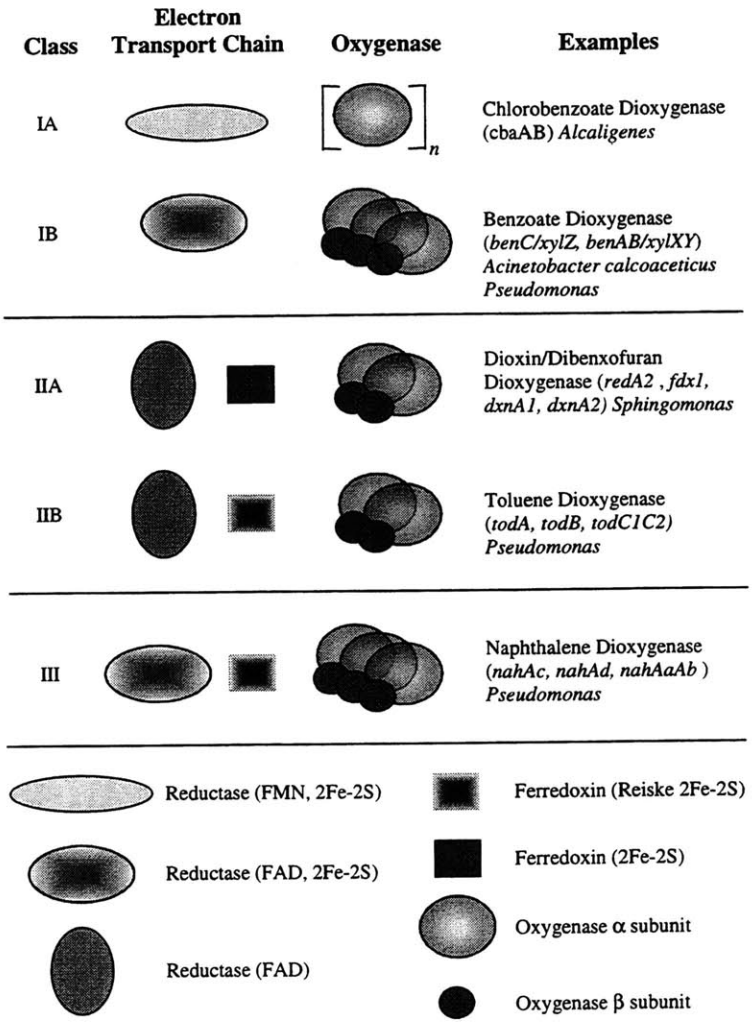


Figure 1-9: Classes of Bacterial Dioxygenases
There are three classes of dioxygenases, both Classes I and II can be subdivided into two groups. The type of electron transport chain and the characteristics of the electron transport proteins define the classes. References: chlorobenzoate dioxygenase (Nakatsu et al. 1995; Nakatsu and Wyndham 1993), benzoate dioxygenase (Harayama and Rekik 1990; Neidle et al. 1991; Neidle et al. 1987; Yamaguchi and Fujisawa 1982), dioxin/dibenzofuran dioxygenase (Armengaud et al. 1998; Armengaud and Timmis 1997; Armengaud and Timmis 1998; Bunz and Cook 1993), toluene dioxygenase (Subramanian et al. 1979; Subramanian et al. 1981; Subramanian et al. 1985; Zylstra and Gibson 1989; Zylstra et al. 1988), and naphthalene dioxygenase (Ensley and Gibson 1983; Haigler and Gibson 1990a; Haigler and Gibson 1990b; Kauppi et al. 1998).

The dioxygenase enzyme mechanism and mode of electron transfer between components are unknown. Recent work has begun to elucidate these mechanisms for some systems. One aspect of the enzyme mechanism is substrate specificity and range; these attributes are highly variable among dioxygenases. Consequently, there are as many different dioxygenases as there are different types of aromatic hydrocarbons. In many cases substrate specificity has been shown, using subunit shuffling, to be controlled by the large subunit (α subunit) of the dioxygenase (Furukawa et al. 1993; Hirose et al. 1994; Suyama 1996; Tan and Cheong 1994). Further analysis of the large subunit and its role in substrate specificity has determined that specific regions of the enzyme can be attributed to controlling substrate selection. These regions are in various locations depending on the dioxygenase, but tend to cluster in the C-terminal portion of the enzyme (Kimura et al. 1997; Mondello 1997; Parales et al. 1998). Furthermore, specific amino acid substitutions have been made that confer new substrate specificity and range to the altered enzyme (Beil et al. 1998; Kimura et al. 1997; Mondello 1997). Other researches, however, have demonstrated that substrate specificity is determined, at least partially, by the small subunit (β subunit) (Butler and Mason 1997; Hirose et al. 1994; Neidle et al. 1991). It seems that each dioxygenase is slightly different in its method of determining substrate specificity.

Electron transfer to the dioxygenase activates the enzyme. A model for electron transfer has been developed from analysis of the crystal structure of the naphthalene dioxygenase from *Pseudomonas putida* (Kauppi et al. 1998). The enzyme is in an $\alpha_3\beta_3$ configuration, contains a Reiske 2Fe-2S center, and a mononuclear iron. Analysis of the structure suggested that an electron is passed from the ferredoxin to the Reiske 2Fe-2S center of the α subunit. This electron is then passed to the mononuclear iron located in the neighboring α subunit, resulting in the activation of the dioxygenase. One particular amino acid, Asp205, acts as an electron bridge between two neighboring α subunits. Replacement of this residue by site directed mutagenesis results in a non-functional electron transport chain and disrupts naphthalene dioxygenase activity (Parales et al. 1999). Site directed mutagenesis of the equivalent amino acid of the toluene dioxygenase from *P. putida* also resulted in a non-functional dioxygenase (Jiang et al. 1996). More research is needed to elucidate the enzyme and electron transfer mechanisms.

Bacterial Diol Dehydrogenases

Dehydrogenases are often categorized into three groups: the long-chain (Persson et al. 1991a), the medium-chain (Persson et al. 1994), and the short-chain dehydrogenases (Jornvall et al. 1995; Krozowski 1994). These classifications are based on the length of the polypeptide. Bacterial diol dehydrogenases are part of the short chain dehydrogenase class. There is only one known exception to this, the benzene dihydrodiol dehydrogenase (*bedD*) from *Pseudomonas putida* ML2 is a medium-chain dehydrogenase (Fong et al. 1996). Other short-chain dehydrogenases include alcohol dehydrogenases and steroid dehydrogenases, and the enzymes are approximately 250 amino acids in length. Unlike the medium

chain varieties, the short chain dehydrogenases do not require metal ions as cofactors (Krozowski 1994). When protein sequences of short chain dehydrogenases are aligned and compared, they contain several strictly conserved residues that define the group including a coenzyme binding site (for NAD^+) and two putative active site residues with one auxiliary residue (Persson et al. 1991b).

Diol dehydrogenases play an important role in aromatic hydrocarbon metabolism. They are generally responsible for the second step of the upper pathway. They prepare the compound for the second oxygenation step by removing the hydrogen atoms from the ring making the compound more reactive and susceptible to oxygen addition by a second dioxygenase. All bacterial diol dehydrogenases studied to date require NAD^+ as a cofactor (Fong et al. 1996; Jeffrey et al. 1975; Khan et al. 1997; Patel and Gibson 1974). The substrate specificities of these enzymes are not well characterized. However, many have been shown to be able to utilize a variety of diol substrates to varying degrees (Rogers and Gibson 1977; Werlen et al. 1996; Keat and Hopper 1978; Patel and Gibson 1974; Sylvestre et al. 1996). A few dehydrogenases have been shown to be stereoselective including the *cis*-biphenyl dihydrodiol dehydrogenase from *Sphingomonas yanoikuyae* B1 (Eaton et al. 1996) and the naphthalene dihydrodiol dehydrogenase from *Pseudomonas putida* (Jeffrey et al. 1975).

Genetics of Aromatic Compound Degradation

There are as many different types of aromatic hydrocarbons as there are different types of dioxygenases. These enzymes have been named unsystematically based on the substrate used to identify the dioxygenase. Upon further analysis of these enzymes it has been found that the substrate range and specificity of these enzymes are highly variable. For example, the naphthalene dioxygenase system from *P. putida* has been shown to have a long and diverse list of possible substrates (Resnick et al. 1996). Consequently the naming of genes for dioxygenase systems is not always reflective of the enzyme's function.

The process of identifying genes for aromatic hydrocarbon metabolism begins with the identification of a strain capable of consuming such compounds as a carbon source. Strains capable of utilizing aromatic compounds as carbon sources have been isolated using classical enrichment protocols. Researchers have used various aromatic compounds, including toluene, naphthalene, benzene, and isopropyl benzene among others, as a carbon source to isolate bacteria. The genes responsible for the metabolism of these compounds have been isolated in a variety of ways. In earlier studies aromatic metabolizing Pseudomonads strains were identified that could utilize naphthalene, toluene, or xylene and then subjected to transposon mutagenesis in order to map the genes. This method was heavily utilized in the analysis of the TOL (toluene utilizing) plasmids and NAH (naphthalene utilizing) plasmid from *Pseudomonas* (Lehrbach et al. 1983; Nakazawa et al. 1980; Yen and Gunsalus 1982). Many of the genes encoding these functions have been cloned using a variety of methods. For example, some of the TOL

plasmid genes from *P. putida* (*xylCMABN*) were cloned by expressing the genes in a heterologous host (*Escherichia coli*) and examining the products formed during aromatic substrate utilization by high pressure liquid chromatography (Harayama et al. 1989). The *nah* genes encoding the utilization of naphthalene from the NAH7 plasmid in *Pseudomonas putida* were cloned in a similar manner (Schell 1983).

Some aromatic degradation operons have been cloned by using an enzyme specific functional assay in a heterologous host. There are two documented colorimetric functional assays that have been used extensively to identify genes. One method identifies ring-hydroxylating dioxygenases, the other identifies ring-cleavage dioxygenases. The identification of ring-hydroxylating dioxygenases takes advantage of the observation that these enzymes have been shown to convert indole to indigo (Ensley et al. 1983); however, in some cases expression of the required enzyme can be host dependent (Kim and Zylstra 1995). For this assay, called the indigo formation assay, indole is added as the substrate in the vapor phase to grown cells. The presence of ring-hydroxylating activity is detected by a color change in the cells ranging from brown to green to blue to purple. This assay has been used to isolate dioxygenase genes and related components from a variety of systems. For example, the isopropyl benzene (*ipb*) genes from *Rhodococcus erythropolis* (Kesseler et al. 1996), the xylene/toluene (*xyl*) genes from *Pseudomonas putida* mt-2 (Keil et al. 1987), a polycyclic aromatic hydrocarbon dioxygenase that can oxygenate naphthalene, phenanthrene, and anthracene from *Pseudomonas fluorescens* (Foght and Westlake 1996), a 2,4-Dinitrotoluene (DNT) dioxygenase from *Burkholderia* (Suen et al. 1996), and phenanthrene/naphthalene dioxygenase from *Comamonas testosteroni* (Goyal and Zylstra 1996) were all isolated using the indigo formation assay.

The second common colorimetric functional assay has been used to identify dioxygenases of the ring-cleavage variety rather than ring-hydroxylating dioxygenases. This assay uses a ring-cleavage substrate that when cleaved produces a colored product. When this substrate is sprayed onto grown colonies that are expressing a ring-cleavage dioxygenase, often the colonies will turn yellow/brown due to the conversion of the compound to a colored product. For example, toluene is converted into 2-hydroxymuconic semialdehyde, which is yellow in color, by the product of the *xylE* gene from *P. putida*. The *xylE* gene was cloned using this colorimetric assay from a genomic library expressed in a heterologous host (Franklin et al. 1981). This method has also been used to clone genes responsible for toluene and xylene metabolism from *Pseudomonas stutzeri* OX1 (Bertoni et al. 1996). Poly-aromatic hydrocarbon ring-cleavage genes have been isolated in a similar fashion. For example, a 3-phenylcatechol dioxygenase from *P. putida* was cloned by screening a cosmid library for clones that turned yellow in the presence of 2,3-dihydroxybiphenyl due to the formation of 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoic acid (Khan et al. 1988). Three different 2,3-dihydroxybiphenyl dioxygenases were identified from *Rhodococcus globerulus* P6 using the same substrate (Asturias and Timmis 1993).

Classical genetic and molecular biology methods have also been used to clone aromatic hydrocarbon degradation operons. The generation of mutants (both chemical and transposon insertion) unable of utilizing an aromatic carbon source and followed by complementation of those mutants with clones from an expression library or cosmid library is a very common methodology. For example, two separate research groups created mutants of the *Rhodococcus erythropolis* IGTS8 that were unable to utilize sulfur using chemical or UV mutagenesis and isolated the *sox/dsz* genes by complementation (Denome 1993; Piddington et al. 1995). Other genes for aromatic hydrocarbon degradation have been isolated in a similar way including the *pca* genes encoding the meta-cleavage pathway of protocatechuate in *Acinetobacter calcoaceticus* (Doten et al. 1987), and the *ben* genes for the benzene lower pathway (also called the benzoate pathway) from *P. putida* (Jeffrey et al. 1992). Furthermore, the availability of sequenced genes has made it possible to use molecular biological techniques such as degenerate polymerase chain reaction (PCR) and other methods based on homology to clone aromatic degradation genes (Gennaro et al. 1997; Hedlund et al. 1999). Sometimes these methods are coupled with prior enzyme purification and amino acid sequencing to provide the necessary sequence information (Eulberg et al. 1997; Strachan et al. 1998).

A novel screening method for mutants defective in toluene degradation developed by Finette and colleagues (Finette et al. 1984) combines a colorimetric assay with classic complementation. The colorimetric assay is based on the redox reactions of the electron transport system and the final electron acceptor, the ring-hydroxylating dioxygenase. Chemically mutagenized cells were screened for strains defective in toluene utilization by the addition of the redox dyes nitro blue tetrazolium (NBT) and 2,3,5-triphenyl-2H-tetrazloium chloride (TTC) in minimal medium. These dyes are different colors depending on their redox state. Wildtype strains will be red due to the reduction of the TTC. Toluene dioxygenase mutants (including the electron transport genes) range from white to light blue depending on the mutation due to the NBT. The *tod* genes for the canonical toluene dioxygenase system in *P. putida* F1 were cloned by complementation using this mutagenesis and screening method (Zylstra et al. 1988).

The subsequent sequencing and analysis of the identified genes has led to some generalizations about the gene organization of aromatic degradation/metabolism systems. There are two overlying themes for the gene organization for operons carrying genes for aromatic compound degradation. Either the genes of the upper and lower pathway appear to be all in one single operon (Zylstra and Gibson 1989), or there are two separate operons that encode the upper and lower pathways separately (Favaro et al. 1996; Franklin et al. 1981; Harayama et al. 1989; Kukor and Olsen 1991; Osborne et al. 1988). Upon closer inspection of the individual operon components, the lower pathways that have been identified are remarkably similar in their gene order (Williams and Sayers 1994). The upper pathways, however, are more variable in their gene order. However, it seems that often the genes are organized in the same order that the enzymes they encode are needed in the pathway. Generally, the components of the

dioxygenase system are clustered together first. They are followed by the genes encoding the diol dehydrogenase and the ring-cleavage dioxygenase (sometimes called the catechol dioxygenase). A generalized schematic of the gene organization of the upper pathway for a single ring aromatic compound is shown in Figure 1-10. These operons can be found on the chromosome (Kiyohara et al. 1994; Polissi et al. 1990; Wright and Olsen 1994), on circular plasmids (Assinder and Williams 1990; Fuenmayor et al. 1998; Williams and Worsey 1976), and linear plasmids (Dabrock et al. 1994). In some cases the plasmid encoded operons are also part of a transposon, some of which are known to actively transpose (Tsuda and Iino 1987) and others that do not (Tsuda and Iino 1990).

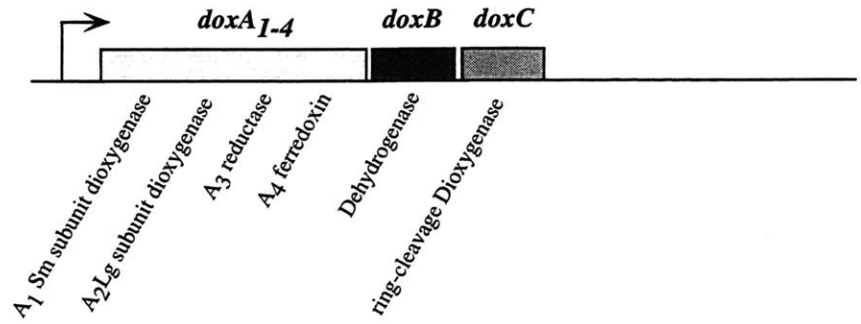


Figure 1-10: General Operon Structure for Upper Pathway Genes
Genes for the upper pathways of aromatic catabolism systems are often organized in a similar manner in different organisms. The diagram shows a general pattern seen in these upper pathway operons.

Actinomycetes

As mentioned above, there is a wide range of organisms capable of utilizing aromatic compounds as carbon sources and in biocatalysis. One such group of resilient bacteria is the actinomycetes. Actinomycetes are Gram-positive, and can be aerobes, anaerobes or facultative anaerobes. They are commonly found in the soil, fresh water, and extreme environments. Actinomycetes are also known to be plant pathogens (El-Raheem et al. 1995; Takeuchi et al. 1996) and human pathogens such as *Corynebacterium diphtheriae* (diphtheria), *Mycobacterium tuberculosis* (tuberculosis), and *Mycobacterium leprae* (leprosy). In biotechnology, actinomycetes are best known for their ability to produce antibiotics (*Streptomcyes*) (Alderson et al. 1993; Baltz 1998), vitamins (Plaut et al. 1974), and enzymes used to produce precursors of pharmaceutical agents (Goodfellow et al. 1984). For example, a precursor to Throbaxane A2 Antagonist, a vasoconstrictor, is produced using enzymes from *Rhodococcus* and *Nocardia globerula* (Patel 1997). Actinomycetes are also used to transform xenobiotics, such as aromatic hydrocarbons, steroids, and pesticides (Peczynska-Czoch and Mordarski 1984).

Actinomycetes are generally described as bacteria with elongated cells or filaments that have the ability to form branching hyphae at some stage of development (Goodfellow et al. 1984). Actinomycetes

have cell walls rich in lipids and a high G+C content (generally > 55%)(Goodfellow and Cross 1984; Williams et al. 1984). The colony morphology can be quite diverse; they can be round, rough, smooth, or mucoidal. Furthermore, the morphology can change over the life cycle of the cells. There are two broad morphological groups of actinomycetes, the sporo- and the nocardioform- actinomycetes. These groups are distinguishable, as their names imply, by the formation of spores by the former and hyphae by the latter. The sporo-actinomycetes comprise the majority of the different species in this group of bacteria. Nocardioform-actinomycetes can also be distinguished from other actinomycetes by the presence of mycolic acids in their cell wall and they tend to have a G+C content around 62% (Williams et al. 1984). Included in the nocardioform actinomycetes are the *Corynebacterium, Mycobacterium, Nocardia,* and *Rhodococcus.*

Rhodococcus

Rhodococci are part of the nocardioform-actinomycetes, many species were previously categorized as *Nocardia.* They are very diverse morphologically and often colored, as their name implies, ranging from buff to orange to red. Changes in morphology of individual cells during the life cycle can cause dramatic effects on the colony morphology (Goodfellow and Cross 1984). Species of Rhodococci are aerobic, non-motile, lysozyme sensitive, and may be acid fast. The genetics and molecular biology of *Rhodococcus* species is not well characterized. Some genetic analysis was conducted with this genus in the late1960's and early 70's, including preliminary genetic mapping via conjugation studies of a *R. erythropolis* strain. However, much of this work was contradictory and therefore inconclusive (Brownell and Denniston 1984). Over the last five years there has been a renewed interest in the genus for their ability to tolerate solvents and catalyze stereospecific enzymatic reactions. Consequently some genetic tools have been developed including plasmids and selectable markers (Dabbs 1990; Quan and Dabbs 1993; Vogt Singer and Finnerty 1988), and transformation protocols (Desomer et al. 1990) to name a few. However, due to the diverse nature of the genus not all of the tools are transferable from one strain to another.

Rhodococci have been found to occupy a variety of environmental niches. There are some species capable of residing in the guts of blood-sucking arthropods (Goodfellow and Cross 1984). Others are pathogenic including *R. equi,* which can be an equine and human pathogen (Prescott 1991), and *R. fascians* which is a plant pathogen (Crespi et al. 1994). As mentioned previously, many isolates of *Rhodococcus* are capable of living in soils contaminated with aromatic compounds (Bell et al. 1998; Kulakova et al. 1996; Warhurst et al. 1994; Warhurst and Fewson 1994). Some Rhodococci have been used in bioconversions. For example, *Rhodococcus* can the desulfurize fuel components (Gray et al. 1996), and a hydrolase from *Rhodococcus* can assist in the synthesis of a β-blocker (Zaks and Dodds 1997). The ability of *Rhodococcus* species to utilize aromatic hydrocarbons and carry out bioconverions makes the species a resource for the bioconversion of indene.

Rhodococcus Strain I24

Rhodococcus strain I24 was isolated as a strain capable of utilizing naphthalene as a sole carbon source from a toluene contaminated aquifer (Buckland et al. 1999; Chartrain et al. 1998). It is a strict aerobe that forms pleomorphic filaments. The strain is also capable of utilizing toluene as a carbon source, but not xylene, biphenyl, or indene. The phylogeny of *Rhodococcus* strain I24 was determined by DNA fingerprint analysis of the tRNA genes and found to cluster with other strains of *Rhodococcus* that are able to oxidize alkylbenzenes (Raymond and Jamison 1968a; Raymond and Jamison 1968b) (Figure 1).

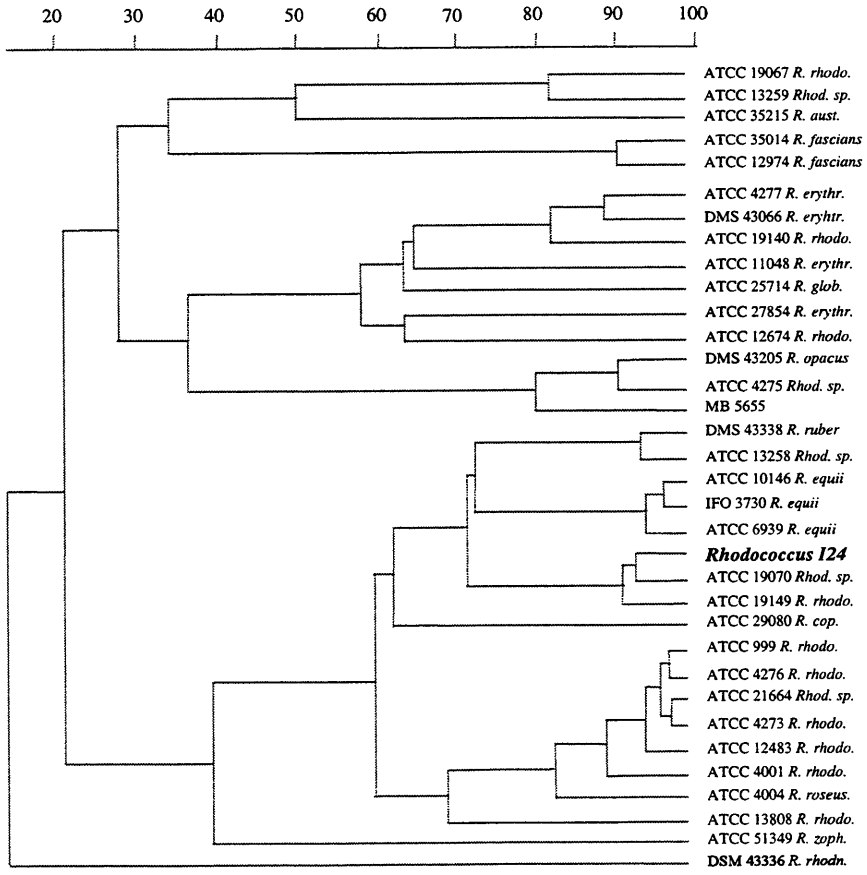


Figure 1-11: Phylogenetic Analysis of *Rhodococcus* strain I24
Figure redrawn from (Chartrain et al. 1998). *Rhodococcus* strain I24 is also known as MB 7209.

Due to its ability to utilize naphthalene and toluene as carbon sources *Rhodococcus* strain I24 was a candidate strain for use in a bioconversion to produce a precursor for the HIV-1 protease inhibitor Crixivan™ as discussed above. Presumably by using the enzymes that allow the utilization of naphthalene and toluene as carbon sources, *Rhodococcus* strain I24 is capable of converting indene to three

enantiomers of indandiol, 1-indenol, 1-indanone, and keto-OH-indan in the presence of a carbon source (Buckland et al. 1999; Chartrain et al. 1998). The production of these compounds was confirmed by NMR analysis (Chartrain et al. 1998). Even though indene is not a carbon source for this strain it can adequately induce the dioxygenase system as evidenced by indene bioconversion. It has been demonstrated in *Pseudomonas* that aromatic degradation systems often need to be induced (Fuenmayor et al. 1998; Mermod et al. 1987). Indene, however, it not likely an optimal inducer for the system since it is not a natural substrate. As a result, Chartrain and colleagues (1998) conducted induction studies with naphthalene and toluene and examined the indene bioconversion product profile. When naphthalene was used to induce the system, *cis*-(1R,2S)-indandiol and *trans*-(1R,2R)-indandiol were produced immediately. When toluene was used as the inducer, *cis*-(1S,2R)-indandiol was the first enantiomer of indandiol produced (Chartrain et al. 1998).

From the bioconversion studies, the induction studies, and the general pathways already established for aromatic hydrocarbon degradation (see above) a pathway for indene bioconversion was proposed for *Rhodococcus* strain I24 (Figure 1-12) (Buckland et al. 1999; Chartrain et al. 1998). In the proposed pathway there may be three different enzyme systems responsible for converting indene into the various 1,2-indandiols in a stereospecific manner. The *Rhodococcus* I24 strain may possess a naphthalene-inducible dioxygenase, a toluene-inducible dioxygenase, and a naphthalene-inducible monooxygenase which all contribute to the oxygenation of indene. The presence of multiple oxygenase enzymes in an organism capable of utilizing a common substrate is not uncommon (Asturias and Timmis 1993; Kosono et al. 1997; Maeda et al. 1995). Additionally, dioxygenases have been found to stereospecifically hydroxylate their substrates (Boyd et al. 1989; Hamberg et al. 1994; Lee and Gibson 1996; Lee et al. 1997; Resnick et al. 1994). Furthermore, oxygenases have been identified that can stereospecifically oxygenate indene (Allen et al. 1997; Gibson et al. 1995; Wackett et al. 1988). A single enzyme complex may be responsible for both the naphthalene-inducible mono- and dioxygenase activities. Earlier work with *Pseudomonas* has demonstrated that naphthalene dioxygenases can act both as dioxygenases and monooxygenases (Gibson et al. 1995; Spain et al. 1989).

As discussed above, oxygenation reactions are generally followed by a dehydrogenation reaction in aromatic hydrocarbon degradation pathways. In *Rhodococcus* strain I24, each of the indandiols produced from indene may be oxidized to keto-OH-indan. These reactions may be carried out by a single dehydrogenase or by separate dehydrogenases specific to each pathway. Overall, the proposed indene bioconversion pathway of *Rhodococcus* strain I24 has multiple oxygenase systems. Each branch of the pathway resembles a portion of the "upper pathway" seen in other aromatic compound degrading bacteria including *Rhodococcus* and *Pseudomonas*.

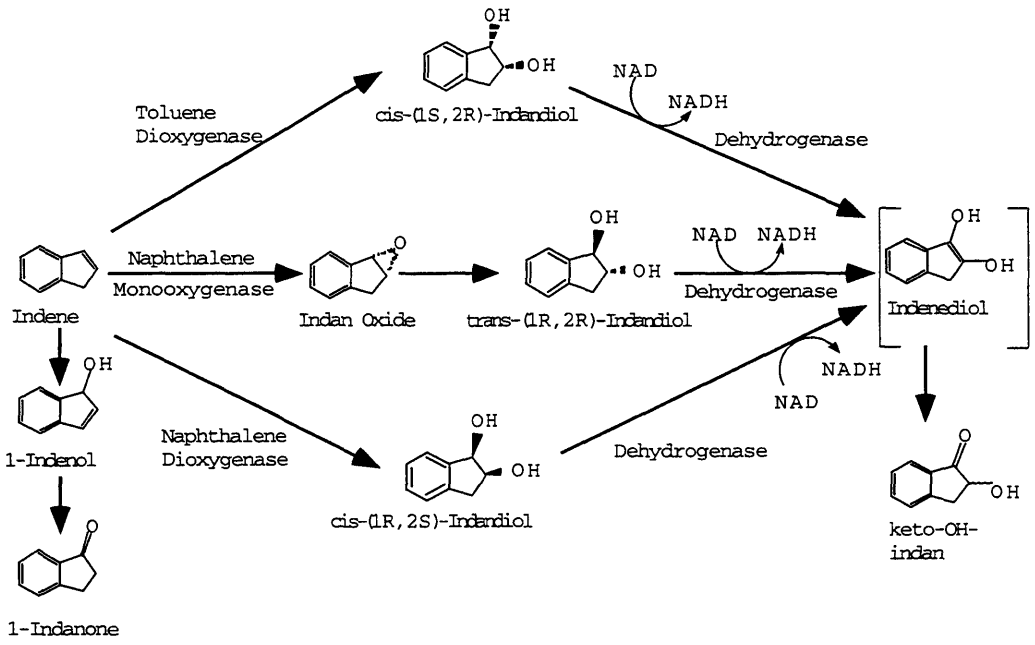


Figure1-12: Proposed Indene Bioconversion pathway from *Rhodococcus* strain I24
This pathway figure was adapted from (Buckland et al. 1999; Chartrain et al. 1998). Indenediol is a putative intermediate not detected by HPLC.

As discussed above, either *trans*-(1R,2R)-indandiol or *cis*-(1S,2R)-indandiol can be used as a precursor for the HIV protease inhibitor indinavir sulfate (Buckland et al. 1999). We are interested in analyzing genetically and biochemically the oxygenation and dehydrogenation steps of indene bioconversion in *Rhodococcus* strain I24. This analysis will provide more insight into the indene bioconversion network to facilitate the development of rational methods to control the production of indandiol in this strain and provide a fundamental understanding of bacterial aromatic compound degradation. More specifically, in this work the identification of a ring-hydroxylating dioxygenase and a diol dehydrogenase from *Rhodococcus* strain I24 are identified. The screening, cloning and initial characterization of the genes are described in chapter two. The characterization of the diol dehydrogenase both genetically and biochemically is described in chapter three.

References

Alderson G, Ritchie DA, Cappellano C, Cool RH, Ivanova NM, Huddleston AS, Flaxman CS, Kristufek V, Lounes A (1993) Physiology and genetics of antibiotic production and resistance. Res Microbiol 144: 665-72

Allen CCR, Boyd DR, Larkin MJ, Reid KA, Sharma ND, Wilson K (1997) Metabolism of Napthalene, 1-Napthol, Indene, and Indole by *Rhodococcus* sp. Strain NCIMB 12038. Appl Environ Microbiol 63: 151-155

Armengaud J, Happe B, Timmis KN (1998) Genetic analysis of dioxin dioxygenase of *Sphingomonas* sp. Strain RW1: catabolic genes dispersed on the genome. J Bacteriol 180: 3954-66

Armengaud J, Timmis KN (1997) Molecular characterization of Fdx1, a putidaredoxin-type [2Fe-2S] ferredoxin able to transfer electrons to the dioxin dioxygenase of *Sphingomonas* sp. RW1. Eur J Biochem 247: 833-42

Armengaud J, Timmis KN (1998) The reductase RedA2 of the multi-component dioxin dioxygenase system of *Sphingomonas* sp. RW1 is related to class-I cytochrome P450-type reductases. Eur J Biochem 253: 437-44

Assinder SJ, Williams PA (1990) The TOL plasmids: determinants of the catabolism of toluene and the xylenes. Adv Microb Physiol 31: 1-69

Asturias JA, Timmis KN (1993) Three Different 2,3-Dihydroxybiphenyl-1,2-Dioxygenase Genes in the Gram-Positive Polychlorobiphenyl-Degrading Bacterium *Rhodococcus globerulus* P6. J Bacteriol 175: 4631-4640

Baltz RH (1998) Genetic manipulation of antibiotic-producing *Streptomyces*. Trends Microbiol 6: 76-83

Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220: 868-71

Batie CJ, Ballou DP, Correll CJ (1991) Phthalate dioxygenase reductase and related flavin-iron-sulfur containing electron transferases. In: Muller F (ed) Chemistry and biochemistry of flavoenzymes. CRC Press, Boca Raton, pp 544-54

Batie CJ, LaHaie E, Ballou DP (1987) Purification and characterization of phthalate oxygenase and phthalate oxygenase reductase from *Pseudomonas cepacia*. J Biol Chem 262: 1510-8

Beil S, Mason JR, Timmis KN, Pieper DH (1998) Identification of chlorobenzene dioxygenase sequence elements involved in dechlorination of 1,2,4,5-tetrachlorobenzene [In Process Citation]. J Bacteriol 180: 5520-8

Bell KS, Philp JC, Aw DW, Christofi N (1998) The genus Rhodococcus. J Appl Microbiol 85: 195-210

Bertoni G, Bolognese F, Galli E, Barbieri P (1996) Cloning of the genes for and characterization of the early stages of toluene and o-xylene catabolism in *Pseudomonas stutzeri* OX1. Appl Environ Microbiol 62: 3704-11

Bhat MA, Vaidyanathan CS (1995) Microbial degradation of halogenated aromatics. In: Singh VP (ed) Biotransformations: Microbial Degradadtion of Health Risk Compounds. Elsevier Science B. V, New York

Boyd DR, McMordie RAS, Sharma ND, Dalton H, Williams P, Jenkins RO (1989) Stereospecific benzylic hydroxylation of bicyclic alkenes by *Pseudomonas putida*: Isolation of (+)-R-1,2-dihydronaphthalene, an arene hydrate of naphthalene from metabolism of 1,2-dihydronaphthalene. J Chem Soc, Chem Commun : 339-340

Brand JM, Cruden DL, Zylstra GJ, Gibson DT (1992) Stereospecific hydroxylation of indan by *Escherichia coli* containing the cloned toluene dioxygenase genes from *Pseudomonas putida* F1 [published erratum appears in Appl Environ Microbiol 1993 Jan;59(1):351]. Appl Environ Microbiol 58: 3407-9

Brownell GH, Denniston K (1984) Genetics of Nocardioform Bacteria. In: Goodfellow M, Modarski M, Williams ST (eds) The biology of Actinomycetes. Academic Press, San Diego, pp 201-228

Buckland B, Drew S, Connors N, Chartrain M, Lee C, Salmon P, Gbewonyo K, Gailliot P, Singhvi R, Olewinski R, Sun W, Reddy J, Zhang J, Zhou W, Jackey B, Goklen K, Junker B, Greasham R (1999) Microbial conversion of Indene to Indandiol, a key intermediate in the synthesis of CRIXIVAN. Metabolic Engineering 1: 63-74

Bunz PV, Cook AM (1993) Dibenzofuran 4,4a-dioxygenase from *Sphingomonas* sp. strain RW1: angular dioxygenation by a three-component enzyme system. J Bacteriol 175: 6467-75

Butler CS, Mason JR (1997) Structure-function Analysis of the Bacterial Aromatic Ring-hydroxylating Dioxyxgenases. In: Poole RK (ed) Advances in Microbial Physiology, vol 38. Academic Press, San Diego, pp 47-84

Chartrain M, Jackey B, Taylor C, Sandford V, Gbewonyo K, Lister L, DiMichele L, Hirsch C, Heimbuch B, Maxwell C, Pascoe D, Buckland B, Greasham R (1998) Bioconversion of indene to *cis*-(1S,2R)-indandiol and *trans*-(1R,2R)-indandiol by *Rhodococcus* species. J Ferment Bioeng 86: 550-558

Crespi M, Vereecke D, Temmerman W, Van Montagu M, Desomer J (1994) The *fas* operon of *Rhodococcus fascians* encodes new genes required for efficient fasciation of host plants. J Bacteriol 176: 2492-501

Dabbs ER, Bhavna Gowan, Susan J. Andersen (1990) Nocardioform Arsenic Resistance Plasmids and Construction of *Rhodococcus* Cloning Vectors. Plasmid 23: 242-247

Dabrock B, Kesseler M, Averhoff B, Gottschalk G (1994) Identification and characterization of a transmissible linear plasmid from *Rhodococcus erythropolis* BD2 that encodes isopropylbenzene and trichloroethene catabolism. Appl Environ Microbiol 60: 853-60

Dagley S (1971) Catabolism of aromatic compounds by micro-organisms. Adv Microb Physiol 6: 1-46

Denome S, Olson ES, Young KD (1993) Identification and cloning of genes involved in specific desulfurization of dibenzothiophene by *Rhodococcus* sp strain IGTS8. Appl Environ Microbiol 59: 2837-2843

Desomer J, Dhaese P, Van Montagu M (1990) Transformation of *Rhodococcus fascians* by High-Voltage Electroporation and Development of *R. fascians* Cloning Vectors. Appl Environ Microbiol 56: 2818-2825

Doten RC, Ngai KL, Mitchell DJ, Ornston LN (1987) Cloning and genetic organization of the *pca* gene cluster from *Acinetobacter calcoaceticus*. J Bacteriol 169: 3168-74

Eaton SL, Resnick SM, Gibson DT (1996) Initial reactions in the oxidation of 1,2-dihydronaphthalene by *Sphingomonas yanoikuyae* strains. Appl Environ Microbiol 62: 4388-94

El-Raheem A, El-Shanshoury R, El-Sououd SM, Awadalla OA, El-Bandy NB (1995) Formation of tomatine in tomato plants infected with *Streptomyces* species and treated with herbicides, correlated with reduction of *Pseudomonas solanacearum* and *Fusarium oxysporum f. sp. lycopersici*. Acta Microbiol Pol 44: 255-66

Ensley BD, Gibson DT (1983) Naphthalene dioxygenase: purification and properties of a terminal oxygenase component. J Bacteriol 155: 505-11

Ensley BD, Ratzkin BJ, Osslund TD, Simon MJ, Wackett LP, Gibson DT (1983) Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. Science 222: 167-9

Eulberg D, Golovleva LA, Schlomann M (1997) Characterization of Catechol Catabolic Genes from *Rhodococcus erythropolis* 1CP. J Bacteriol 179: 370-381

Favaro R, Bernasconi C, Passini N, Bertoni G, Bestetti G, Galli E, Deho G (1996) Organisation of the *tmb* catabolic operons of *Pseudomonas putida* TMB and evolutionary relationship with the *xyl* operons of the TOL plasmid pWW0. Gene 182: 189-93

Feigelson P (1976) Pseudomonad and hepatic L-tryptophan 2, 3-dioxygenase. Adv Exp Med Biol 74: 358-62

Fetzner S (1998) Bacterial dehalogenation. Appl Microbiol Biotechnol 50: 633-57

Fetzner S, Muller R, Lingens F (1992) Purification and some properties of 2-halobenzoate 1,2-dioxygenase, a two-component enzyme system from *Pseudomonas cepacia* 2CBS. J Bacteriol 174: 279-90

Finette BA, Subramanian V, Gibson DT (1984) Isolation and characterization of *Pseudomonas putida* PpF1 mutants defective in the toluene dioxygenase enzyme system. J Bacteriol 160: 1003-9

Foght JM, Westlake DW (1996) Transposon and spontaneous deletion mutants of plasmid-borne genes encoding polycyclic aromatic hydrocarbon degradation by a strain of *Pseudomonas fluorescens*. Biodegradation 7: 353-66

Fong KP, Goh CB, Tan HM (1996) Characterization and expression of the plasmid-borne *bedD* gene from *Pseudomonas putida* ML2, which codes for a NAD+-dependent cis-benzene dihydrodiol dehydrogenase. J Bacteriol 178: 5592-601

Frankel AD, Young JA (1998) HIV-1: fifteen proteins and an RNA. Annu Rev Biochem 67: 1-25

Franklin FC, Bagdasarian M, Bagdasarian MM, Timmis KN (1981) Molecular and functional analysis of the TOL plasmid pWWO from *Pseudomonas putida* and cloning of genes for the entire regulated aromatic ring meta cleavage pathway. Proc Natl Acad Sci U S A 78: 7458-62

Fuenmayor SL, Wild M, Boyes AL, Williams PA (1998) A gene cluster encoding steps in conversion of naphthalene to gentisate in *Pseudomonas* sp. strain U2. J Bacteriol 180: 2522-30

Furukawa K, Hirose J, Suyama A, Zaiki T, Hayashida S (1993) Gene components responsible for discrete substrate specificity in the metabolism of biphenyl (*bph* operon) and toluene (*tod* operon). J Bacteriol 175: 5224-32

Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, Mann D, Sidhu GD, Stahl RE, Zolla-Pazner S, Leibowitch J, Popovic M (1983) Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science 220: 865-7·

Geary PJ, Mason JR, Joannou CL (1990) Benzene dioxygenase from *Pseudomonas putida* ML2 (NCIB 12190). Methods Enzymol 188: 52-60

Gennaro PD, Galli E, Albini G, Pelizzoni F, Sello G, Bestetti G (1997) Production of substituted naphthalene dihydrodiols by engineered *Escherichia coli* containing the cloned naphthalene 1,2-dioxygenase gene from *Pseudomonas fluorescens* N3. Res Microbiol 148: 355-64

Gibson DT, Resnick SM, Lee K, Brand JM, Torok DS, Wackett LP, Schocken MJ, Haigler BE (1995) Desaturation, Dioxygenation, and Monooxygenation Reactions Catalyzed by Napthalene Dioxygenase from *Pseudomonas* sp. Strain 9816-4. J Bacteriol 177: 2615-2621

Gibson DT, Subramanian V (1984) Microbial Degradation of Aromatic Hydrocarbons. In: Gibson DT (ed) Microbial Degradation of Organic Compounds, vol 13. Macel Dekker, Inc., New York, pp 181-252

Glazer AN, Nikaido H (1995) Microbial Biotechnology. W. H. Freeman and Company, New Tork

Goodfellow M, Cross T (1984) Classification [of Actinomycetes]. In: Goodfellow M, Mordarski M, Williams ST (eds) The Biology of Actinomycetes. Academic Press, London, pp 7-164

Goodfellow M, Mordarski M, Williams ST (1984) Introduction to and Importance of Actinomycetes. In: Goodfellow M, Mordarski M, Williams ST (eds) The Biology of Actinomycetes. Academic Press, London, pp 1-6

Goyal AK, Zylstra GJ (1996) Molecular cloning of novel genes for polycyclic aromatic hydrocarbon degradation from *Comamonas testosteroni* GZ39. Appl Environ Microbiol 62: 230-6

Gray KA, Pogrebinsky OS, Mrachko GT, Xi L, Monticello DJ, Squires CH (1996) Molecular mechanisms of biocatalytic desulfurization of fossil fuels [In Process Citation]. Nat Biotechnol 14: 1705-9

Grund E, Schmitz A, Fiedler J, Gartemann K-H (1995) Aerobic biodegradation of polycyclic and halogenated aromatic compounds. In: Singh VP (ed) Biotransformations: Microbial Degradation of Health Risk Compounds. Elsevier Science B.V, New York

Haddock JD, Gibson DT (1995) Purification and characterization of the oxygenase component of biphenyl 2,3-dioxygenase from *Pseudomonas* sp. strain LB400 [published erratum appears in J Bacteriol 1996 Apr;178(7):258]. J Bacteriol 177: 5834-9

Hahn BH, Shaw GM, Arya SK, Popovic M, Gallo RC, Wong-Staal F (1984) Molecular cloning and characterization of the HTLV-III virus associated with AIDS. Nature 312: 166-9

Haigler BE, Gibson DT (1990a) Purification and properties of ferredoxinNAP, a component of naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816. J Bacteriol 172: 465-8

Haigler BE, Gibson DT (1990b) Purification and properties of NADH-ferredoxinNAP reductase, a component of naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816. J Bacteriol 172: 457-64

Hamberg M, Zhang LY, Brodowsky ID, Oliw EH (1994) Sequential oxygenation of linoleic acid in the fungus *Gaeumannomyces graminis*: stereochemistry of dioxygenase and hydroperoxide isomerase reactions. Arch Biochem Biophys 309: 77-80

Hammer SM, Squires KE, Hughes MD, Grimes JM, Demeter LM, Currier JS, Eron JJ, Jr., Feinberg JE, Balfour HH, Jr., Deyton LR, Chodakewitz JA, Fischl MA (1997) A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team [see comments]. N Engl J Med 337: 725-33

Harayama S, Kok M, Neidle EL (1992) Functional and evolutionary relationships among diverse oxygenases. Annu Rev Microbiol 46: 565-601

Harayama S, Rekik M (1990) The meta cleavage operon of TOL degradative plasmid pWW0 comprises 13 genes. Mol Gen Genet 221: 113-20

Harayama S, Rekik M, Wubbolts M, Rose K, Leppik RA, Timmis KN (1989) Characterization of five genes in the upper-pathway operon of TOL plasmid pWW0 from *Pseudomonas putida* and identification of the gene products. J Bacteriol 171: 5048-55

Hayaishi O (1962) Chapter 1: History and Scope. In: Hayaishi O (ed) Oxygenases. Academic Press, New York

Hedlund BP, Geiselbrecht AD, Bair TJ, Staley JT (1999) Polycyclic aromatic hydrocarbon degradation by a new marine bacterium, *Neptunomonas naphthovorans* gen. nov., sp. nov. Appl Environ Microbiol 65: 251-9

Hirose J, Suyama A, Hayashida S, Furukawa K (1994) Construction of hybrid biphenyl (*bph*) and toluene (*tod*) genes for functional analysis of aromatic ring dioxygenases. Gene 138: 27-33

Jeffrey AM, Yeh HJ, Jerina DM, Patel TR, Davey JF, Gibson DT (1975) Initial reactions in the oxidation of naphthalene by *Pseudomonas putida*. Biochemistry 14: 575-84

Jeffrey WH, Cuskey SM, Chapman PJ, Resnick S, Olsen RH (1992) Characterization of *Pseudomonas putida* mutants unable to catabolize benzoate: cloning and characterization of *Pseudomonas* genes involved in benzoate catabolism and isolation of a chromosomal DNA fragment able to substitute for *xylS* in activation of the TOL lower-pathway promoter. J Bacteriol 174: 4986-96

Jiang H, Parales RE, Lynch NA, Gibson DT (1996) Site-directed mutagenesis of conserved amino acids in the alpha subunit of toluene dioxygense: potential monobuclear non-heme iron coordination sites. J Bacteriol 178: 3133-3139

Jornvall H, Persson B, Krook M, Atrian S, Gonzalez-Duarte R, Jeffery J, Ghosh D (1995) Short-chain dehydrogenases/reductases (SDR). Biochemistry 34: 6003-13

Katz RA, Skalka AM (1994) The retroviral enzymes. Annu Rev Biochem 63: 133-73

Kauppi B, Lee K, Carredano E, Parales RE, Gibson DT, Eklund H, Ramaswamy S (1998) Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2- dioxygenase. Structure 6: 571-86

Keat MJ, Hopper DJ (1978) The aromatic alcohol dehydrogenases in *Pseudomonas putida* N.C.I.B. 9869 grown on 3,5-xylenol and p-cresol. Biochem J 175: 659-67

Keil H, Saint CM, Williams PA (1987) Gene organization of the first catabolic operon of TOL plasmid pWW53: production of indigo by the *xylA* gene product. J Bacteriol 169: 764-70

Kesseler M, Dabbs ER, Averhoff B, Gottschalk G (1996) Studies on the isopropylbenzene 2,3-dioxygenase and the 3-isopropylcatechol 2,3-dioxygenase genes encoded by the linear plasmid of *Rhodococcus erythropolis* BD2. Microbiology 142: 3241-3251

Khan A, Tewari R, Walia S (1988) Molecular cloning of 3-phenylcatechol dioxygenase involved in the catabolic pathway of chlorinated biphenyl from *Pseudomonas putida* and its expression in *Escherichia coli*. Appl Environ Microbiol 54: 2664-71

Khan A, Walia S (1989) Cloning of bacterial genes specifying degradation of 4-chlorobiphenyl from *Pseudomonas putida* OU83. Appl Environ Microbiol 55: 798-805

Khan AA, Wang RF, Nawaz MS, Cerniglia CE (1997) Nucleotide sequence of the gene encoding cis-biphenyl dihydrodiol dehydrogenase (*bphB*) and the expression of an active recombinant His- tagged *bphB* gene product from a PCB degrading bacterium, *Pseudomonas putida* OU83. FEMS Microbiol Lett 154: 317-24

Kim E, Zylstra GJ (1995) Molecular and biochemical characterization of two meta-cleavage dioxygenases involved in biphenyl and m-xylene degradation by *Beijerinckia* sp. strain B1. J Bacteriol 177: 3095-103

Kimura N, Nishi A, Goto M, Furukawa K (1997) Functional analyses of a variety of chimeric dioxygenases constructed from two biphenyl dioxygenases that are similar structurally but different functionally. J Bacteriol 179: 3936-43

Kiyohara H, Torigoe S, Kaida N, Asaki T, Iida T, Hayashi H, Takizawa N (1994) Cloning and characterization of a chromosomal gene cluster, *pah*, that encodes the upper pathway for phenanthrene and naphthalene utilization by *Pseudomonas putida* OUS82. J Bacteriol 176: 2439-43

Klotman ME, Wong-Staal F (1991) Human Immunodeficiency Virus (HIV) Gene Structure and Genetic Diversity. In: Gallo RC, Jay G (eds) The Human Retroviruses. Academic Press, San Diego

Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RA, Scolnick EM, Sigal IS (1988) Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci U S A 85: 4686-90

Kosono S, Maeda M, Fuji F, Arai H, Kudo T (1997) Three of the Seven *bphC* Genes of *Rhodococcus erythropolis* TA421, Isolated from a Termite Ecosystem, Are Located on an Indigenous Plasmid Associated with Biphenyl Degradation. Appl Environ Microbiol 63: 3282-3285

Krozowski Z (1994) The short-chain alcohol dehydrogenase superfamily: variations on a common theme. J Steroid Biochem Mol Biol 51: 125-30

Kukor JJ, Olsen RH (1991) Genetic organization and regulation of a meta cleavage pathway for catechols produced from catabolism of toluene, benzene, phenol, and cresols by *Pseudomonas pickettii* PKO1. J Bacteriol 173: 4587-94

Kulakova AN, Reid KA, Larkin MJ, Allen CC, Kulakov LA (1996) Isolation of *Rhodococcus rhodochrous* NCIMB13064 derivatives with new biodegradative abilities. FEMS Microbiol Lett 145: 227-31

Lee K, Gibson DT (1996) Stereospecific dihydroxylation of the styrene vinyl group by purified naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816-4. J Bacteriol 178: 3353-6

Lee K, Resnick SM, Gibson DT (1997) Stereospecific Oxidation of (R)- and (S)-1-Indanol by Naphthalene Dioxygenase from *Pseudomonas* sp. Strain NCIB 9816-4. Appl Environ Microbiol 63: 2067-2070

Lehning A, Fock U, Wittich RM, Timmis KN, Pieper DH (1997) Metabolism of chlorotoluenes by *Burkholderia* sp. strain PS12 and toluene dioxygenase of *Pseudomonas putida* F1: Evidence for monooygenation by toluene and chlorobenzene dioxygenases. Appl Environ Microbiol 63: 1974-1979

Lehrbach PR, Jeenes DJ, Broda P (1983) Characterization by molecular cloning of insertion mutants in TOL catabolic functions. Plasmid 9: 112-25

Lei B, Tu SC (1996) Gene overexpression, purification, and identification of a desulfurization enzyme from *Rhodococcus* sp. strain IGTS8 as a sulfide/sulfoxide monooxygenase. J Bacteriol 178: 5699-705

Locher HH, Leisinger T, Cook AM (1991) 4-Sulphobenzoate 3,4-dioxygenase. Purification and properties of a desulphonative two-component enzyme system from *Comamonas testosteroni* T-2. Biochem J 274: 833-42

Luciw PA, Potter SJ, Steimer K, Dina D, Levy JA (1984) Molecular cloning of AIDS-associated retrovirus. Nature 312: 760-3

Maeda M, Chung SY, Song E, Kudo T (1995) Multiple genes encoding 2,3-dihydroxybiphenyl 1,2-dioxygenase in the gram-positive polychlorinated biphenyl-degrading bacterium *Rhodococcus erythropolis* TA421, isolated from a termite ecosystem. Appl Environ Microbiol 61: 549-55

Maltseva OV, Solyanikova IP, Golovleva LA (1994) Chlorocatechol 1,2-dioxygenase from *Rhodococcus erythropolis* 1CP. Kinetic and immunochemical comparison with analogous enzymes from gram-negative strains. Eur J Biochem 226: 1053-61

Mason JR (1988) Oxygenase catalyzed hydroxylation of aromatic compounds: simple chemistry by complex enzymes. Int Ind Biotechnol 8: 19-24

Mason JR, Cammack R (1992) The electron-transport proteins of hydroxylating bacterial dioxygenases. Annu Rev Microbiol 46: 277-305

McQuade TJ, Tomasselli AG, Liu L, Karacostas V, Moss B, Sawyer TK, Heinrikson RL, Tarpley WG (1990) A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. Science 247: 454-6

Mermod N, Ramos JL, Bairoch A, Timmis KN (1987) The *xylS* gene positive regulator of TOL plasmid pWWO: identification, sequence analysis and overproduction leading to constitutive expression of meta cleavage operon. Mol Gen Genet 207: 349-54

Mondello FJ, Michael P. Turcich, John H. Lobos, Bruce D. Erickson (1997) Identification and Modification of Biphenyl Dioxygenase Sequences That Determine the Specificity of Polychlorinated Biphenyl Degradation. *Applied and Environmental Microbiology* 63: 3096-3103

Mueller JG, Devereux R, Santavy DL, Lantz SE, Willis SG, Pritchard PH (1997) Phylogenetic and physiological comparisons of PAH-degrading bacteria from geographically diverse soils. Antonie Van Leeuwenhoek 71: 329-43

Nakatsu CH, Straus NA, Wyndham RC (1995) The nucleotide sequence of the Tn5271 3-chlorobenzoate 3,4-dioxygenase genes (cbaAB) unites the class IA oxygenases in a single lineage. Microbiology 141: 485-95

Nakatsu CH, Wyndham RC (1993) Cloning and expression of the transposable chlorobenzoate-3,4-dioxygenase genes of *Alcaligenes* sp. strain BR60. Appl Environ Microbiol 59: 3625-33

Nakazawa T, Inouye S, Nakazawa A (1980) Physical and functional mapping of RP4-TOL plasmid recombinants: analysis of insertion and deletion mutants. J Bacteriol 144: 222-31

Neidle EL, Hartnett C, Ornston LN, Bairoch A, Rekik M, Harayama S (1991) Nucleotide sequences of the *Acinetobacter calcoaceticus benABC* genes for benzoate 1,2-dioxygenase reveal evolutionary relationships among multicomponent oxygenases. J Bacteriol 173: 5385-95

Neidle EL, Shapiro MK, Ornston LN (1987) Cloning and expression in *Escherichia coli* of *Acinetobacter calcoaceticus* genes for benzoate degradation. J Bacteriol 169: 5496-503

Oldfield C, Pogrebinsky O, Simmonds J, Olson ES, Kulpa CF (1997) Elucidation of the metabolic pathway for dibenzothiophene desulphurization by *Rhodococcus* sp. strain IGTS8 (ATCC 53968). Microbiology 143: 2961-73

Osborne DJ, Pickup RW, Williams PA (1988) The presence of two complete homologous meta pathway operons on TOL plasmid pWW53. J Gen Microbiol 134: 2965-75

Parales RE, Emig MD, Lynch NA, Gibson DT (1998) Substrate specificities of hybrid naphthalene and 2,4-dinitrotoluene dioxygenase enzyme systems. J Bacteriol 180: 2337-44

Parales RE, Parales JV, Gibson DT (1999) Aspartate 205 in the catalytic domain of naphthalene dioxygenase is essential for activity. J Bacteriol 181: 1831-1837

Patel RN (1997) Stereoselcetive Biotransformations in Syntheis of Some Pharmaceutical Intermediates. In: Neidleman SL, Laskin AI (eds) Advacnec in Applied Micribiology, vol 43. Academic Press, San Diego, pp 91-140

Patel TR, Gibson DT (1974) Purification and properties of (+)-cis-naphthalene dihydrodiol dehydrogenase of *Pseudomonas putida*. J Bacteriol 119: 879-88

Peczynska-Czoch W, Mordarski M (1984) Transformation of Xenobiotics. In: Goodfellow M, Mordarski M, Williams ST (eds) The Biology of Actinomycetes. Academic Press, London, pp 287-336

Persson B, Jeffery J, Jornvall H (1991a) Different segment similarities in long-chain dehydrogenases. Biochem Biophys Res Commun 177: 218-23

Persson B, Krook M, Jornvall H (1991b) Characteristics of short-chain alcohol dehydrogenases and related enzymes. Eur J Biochem 200: 537-43

Persson B, Zigler JS, Jr., Jornvall H (1994) A super-family of medium-chain dehydrogenases/reductases (MDR). Sub- lines including zeta-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductase enoyl reductases, VAT-1 and other proteins. Eur J Biochem 226: 15-22

Piddington CS, Kovacevich BR, Rambosek J (1995) Sequence and molecular characterization of a DNA region encoding the dibenzothiophene desulfurization operon of *Rhodococcus* sp. strain IGTS8. Appl Environ Microbiol 61: 468-75

Plaut GW, Smith CM, Alworth WL (1974) Biosynthesis of water-soluble vitamins. Annu Rev Biochem 43: 899-922

Polissi A, Bestetti G, Bertoni G, Galli E, Deho G (1990) Genetic analysis of chromosomal operons involved in degradation of aromatic hydrocarbons in *Pseudomonas putida* TMB. J Bacteriol 172: 6355-62

Prescott JF (1991) *Rhodococcus equi*: an animal and human pathogen. Clin Microbiol Rev 4: 20-34

Primrose SB (1991) Molecular Biotechnology, 2nd edn. Balckwell Scientific Publications, Cambridge

Quan S, Dabbs ER (1993) Nocardioform arsenic resistance plasmid characterization and improved *Rhodococcus* cloning vectors. Plasmid 29: 74-9

Raymond RL, Jamison VW (1968a) ATCC #19149. . American Type Culture Collection (ATCC), Manassas, VA

Raymond RL, Jamison VW (1968b) ATCC # 19070. . American Type Culture Collection (ATCC), Manassas, VA

Reider PJ (1997) Advances in AIDS chemotherapy: The assymmetric synthesis of CRIXIVAN. Chimia 51: 381-386

Resnick SM, Lee K, Gibson DT (1996) Diverse reactions catalyzed by naphthalene dioxygenase from *Pseudomonas* sp strain NCIB 9816. J Ind Micro 17: 438-57

Resnick SM, Torok DS, Lee K, Brand JM, Gibson DT (1994) Regiospecific and Stereoselective Hydroxylation of 1-Indanone and 2-Indanone by Napthalene Dioxygenase and Toluene Dioxygenase. Appl Environ Microbiol 60: 3323-3328

Rogers JE, Gibson DT (1977) Purification and properties of cis-toluene dihydrodiol dehydrogenase from *Pseudomonas putida.* J Bacteriol 130: 1117-24

Rutschmann O, Hirschel B (1997) Antiretroviral therapy: a guide to the most important trials. Schweiz Med Wochenschr 127: 436-43

Sanseverino J, Applegate BM, King JM, Sayler GS (1993) Plasmid-mediated mineralization of naphthalene, phenanthrene, and anthracene. Appl Environ Microbiol 59: 1931-7

Sarngadharan MG, DeVico AL, Bruch L, Schupbach J, Gallo RC (1984) HTLV-III: the etiologic agent of AIDS. Princess Takamatsu Symp 15: 301-8

Sauber K, Frohner C, Rosenberg G, Eberspacher J, Lingens F (1977) Purification and properties of pyrazon dioxygenase from pyrazon- degrading bacteria. Eur J Biochem 74: 89-97

Schell MA (1983) Cloning and expression in *Escherichia coli* of the naphthalene degradation genes from plasmid NAH7. J Bacteriol 153: 822-9

Seelmeier S, Schmidt H, Turk V, von der Helm K (1988) Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. Proc Natl Acad Sci U S A 85: 6612-6

Simon MJ, Osslund TD, Saunders R, Ensley BD, Suggs S, Harcourt A, Suen WC, Cruden DL, Gibson DT, Zylstra GJ (1993) Sequences of genes encoding naphthalene dioxygenase in *Pseudomonas putida* strains G7 and NCIB 9816-4. Gene 127: 31-7

Spain JC (1995) Biodegradation of nitroaromatic compounds. Annu Rev Microbiol 49: 523-55

Spain JC, Zylstra GJ, Blake CK, Gibson DT (1989) Monohydroxylation of phenol and 2,5-dichlorophenol by toluene dioxygenase in *Pseudomonas putida* F1 [published erratum appears in Appl Environ Microbiol 1991 May;57(5):1579]. Appl Environ Microbiol 55: 2648-52

Strachan PD, Freer AA, Fewson CA (1998) Purification and characterization of catechol 1,2-dioxygenase from *Rhodococcus rhodochrous* NCIMB 13259 and cloning and sequencing of its *catA* gene. Biochem J 333: 741-7

Subramanian V, Liu TN, Yeh WK, Gibson DT (1979) Toluene dioxygenase: purification of an iron-sulfur protein by affinity chromatography. Biochem Biophys Res Commun 91: 1131-9

Subramanian V, Liu TN, Yeh WK, Narro M, Gibson DT (1981) Purification and properties of NADH-ferredoxinTOL reductase. A component of toluene dioxygenase from *Pseudomonas putida*. J Biol Chem 256: 2723-30

Subramanian V, Liu TN, Yeh WK, Serdar CM, Wackett LP, Gibson DT (1985) Purification and properties of ferredoxinTOL. A component of toluene dioxygenase from *Pseudomonas putida* F1. J Biol Chem 260: 2355-63

Suen WC, Haigler BE, Spain JC (1996) 2,4-Dinitrotoluene dioxygenase from *Burkholderia* sp. strain DNT: similarity to naphthalene dioxygenase. J Bacteriol 178: 4926-34

Suyama A, Ryo Iwakiri, Nobutada Kimura, Akito Nishi, Kanji Nakamura, Kensuke Furukawa (1996) Engineering Hybrid Pseudomonads Capable of Utilizing a Wide Range of Aromatic Hydrocarbons and of Efficient Degradation of Trichloroethylene. Journal of Bacteriology 178: 4039-4046

Sylvestre M, Hurtubise Y, Barriault D, Bergeron J, Ahmad D (1996) Characterization of active recombinant 2,3-dihydro-2,3- dihydroxybiphenyl dehydrogenase from *Comamonas testosteroni* B-356 and sequence of the encoding gene (*bphB*). Appl Environ Microbiol 62: 2710-5

Takeuchi T, Sawada H, Tanaka F, Matsuda I (1996) Phylogenetic analysis of *Streptomyces* spp. causing potato scab based on 16S rRNA sequences. Int J Syst Bacteriol 46: 476-9

Tan H, Cheong C (1994) Substitution of the ISP α subunit of biphenyl dioxygenase from *Pseudomonas* results in a modification of enzyme activity. Biochem Biophys Res Commun 204: 912-917

Tsuda M, Iino T (1987) Genetic analysis of a transposon carrying toluene degrading genes on a TOL plasmid pWW0. Mol Gen Genet 210: 270-6

Tsuda M, Iino T (1990) Naphthalene degrading genes on plasmid NAH7 are on a defective transposon. Mol Gen Genet 223: 33-9

Vacca JP, Condra JH (1997) Clinically effective HIV-1 protease inhibitors. Drug Discovery Today 2: 261-272

Vogt Singer ME, Finnerty WR (1988) Construction of an *Escherichia coli-Rhodococcus* Shuttle Vector and Plasmid Transformation in *Rhodococcus* spp. J Bacteriol 170: 638-645

Wackett LP, Kwart LD, Gibson DT (1988) Benzylic monooxygenation catalyzed by toluene dioxygenase from *Pseudomonas putida*. Biochemistry 27: 1360-7

Warhurst AM, Clarke KF, Hill RA, Holt RA, Fewson CA (1994) Metabolism of styrene by *Rhodococcus rhodochrous* NCIMB 13259. Appl Environ Microbiol 60: 1137-45

Warhurst AM, Fewson CA (1994) Biotransformations catalyzed by the genus *Rhodococcus*. Crit Rev Biotechnol 14: 29-73

Werlen C, Kohler HP, van der Meer JR (1996) The broad substrate chlorobenzene dioxygenase and cis-chlorobenzene dihydrodiol dehydrogenase of *Pseudomonas* sp. strain P51 are linked evolutionarily to the enzymes for benzene and toluene degradation. J Biol Chem 271: 4009-16

Williams PA, Sayers JR (1994) The evolution of pathways for aromatic hydrocarbon oxidation in *Pseudomonas*. Biodegradation 5: 195-217

Williams PA, Worsey MJ (1976) Ubiquity of plasmids in coding for toluene and xylene metabolism in soil bacteria: evidence for the existence of new TOL plasmids. J Bacteriol 125: 818-28

Williams ST, Sharpe ME, Holt JG (1984) Bergey's Manual of Systematic Bacteriology. , vol 4. Williams and Wilkins, Baltimore

Wlodawer A, Vondrasek J (1998) Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. Annu Rev Biophys Biomol Struct 27: 249-84

Wright A, Olsen RH (1994) Self-mobilization and organization of the genes encoding the toluene metabolic pathway of *Pseudomonas mendocina* KR1. Appl Environ Microbiol 60: 235-42

Yamaguchi M, Fujisawa H (1982) Subunit structure of oxygenase component in benzoate-1,2-dioxygenase system from *Pseudomonas arvilla* C-1. J Biol Chem 257: 12497-502

Yen KM, Gunsalus IC (1982) Plasmid gene organization: naphthalene/salicylate oxidation. Proc Natl Acad Sci U S A 79: 874-8

Zaks A, Dodds DR (1997) Application of biocatalysis and biotransformations to the synthesis of pharmaceuticals. Drug Discovery Today 2: 513-531

Zimmermann TP, Robbins KT, Werlen J, Hoeks FWJMM (1997) Bio-transformation in the Production of L-Carnitine. In: Collins AN, Sheldrake GN, Crosby J (eds) Chirality in Industry II: Developments in Manufacturing and Applications of Optically Active Compounds. John Wiley & Sons, Inc., New York, pp 287-305

Zylstra GJ (1994) Molecular Analysis of Aromatic Hydrocarbon Degradation. In: Garte SJ (ed) Molecular Environmental Biology. Lewis Publishers, Boca Raton, pp 83-115

Zylstra GJ, Gibson DT (1989) Toluene degradation by *Pseudomonas putida* F1. Nucleotide sequence of the *todC1C2BADE* genes and their expression in *Escherichia coli*. J Biol Chem 264: 14940-6

Zylstra GJ, McCombie WR, Gibson DT, Finette BA (1988) Toluene degradation by *Pseudomonas putida* F1: genetic organization of the tod operon. Appl Environ Microbiol 54: 1498-503

A portion of the following chapter is included in a publication. The reference for this publication is:

Treadway SL, Yanagimachi K, Lankenau E, Lessard PA, Stephanopoulos G, Sinskey AJ (1999) Isolation and characterization of indene bioconversion genes from *Rhodococcus* strain I24. Appl Microbiol Biotechnol **In press.**

Chapter Two: Isolation and Characterization of Indene Bioconversion Genes from *Rhodococcus* strain I24

Abstract

Rhodococcus strain I24 is able to convert indene into indandiol via the actions of at least two dioxygenase systems and a putative monooxygenase system. We have identified a cosmid clone from I24 genomic DNA that is able to confer the ability to convert indene to indandiol upon *Rhodococcus erythropolis* SQ1, a strain that normally can not convert or metabolize indene. HPLC analysis revealed that the transformed SQ1 strain produces *cis*-(1R, 2S)-indandiol suggesting that the cosmid clone encodes a naphthalene-type dioxygenase. DNA sequence analysis of a portion of this clone confirmed the presence of genes for the dioxygenase as well as genes encoding a dehydrogenase and putative aldolase. These genes will be useful for manipulating indene bioconversion in *Rhodococcus* strain I24.

Introduction

The ability of many bacteria to metabolize aromatic compounds makes them useful tools to synthesize enantiopure precursors of biologically active products (Collins et al. 1997). Common precursors are diols (Carless 1992) that can be produced in oxygenation reaction carried out by multicomponent dioxygenase systems (Mason and Cammack 1992), or produced by successive monooxygenation reactions. Dioxygenation is the initial reaction in the breakdown of many aromatic compounds and is generally followed by a dehydrogenation reaction (Grund et al. 1995; Zylstra 1994). There are many examples of such aromatic degradation pathways in both Gram negative and Gram positive bacteria including *Pseudomonas* and *Rhodococcus* (Allen et al. 1997; Masai et al. 1995; Zylstra et al. 1989).

Rhodococcus strain I24, isolated from soil contaminated with aromatic compounds, is capable of metabolizing naphthalene and toluene as sole carbon sources (Chartrain et al. 1998). Although this strain can not use indene as a sole carbon source, it can oxygenate indene in the presence of a carbon source. It is hypothesized that the same oxygenase systems that allow catabolism of naphthalene and toluene allow *Rhodococcus* strain I24 to oxygenate indene. Indene is converted to a variety of indandiols, two of which can be utilized as precursors in the synthesis of the HIV protease inhibitor indinavir sulfate (Buckland et al. 1999). A pathway for indene bioconversion in this strain of *Rhodococcus* has been proposed based on the analysis of indene breakdown in the presence and absence of naphthalene and toluene as inducers (Figure 2-1)(Buckland et al. 1999; Chartrain et al. 1998).

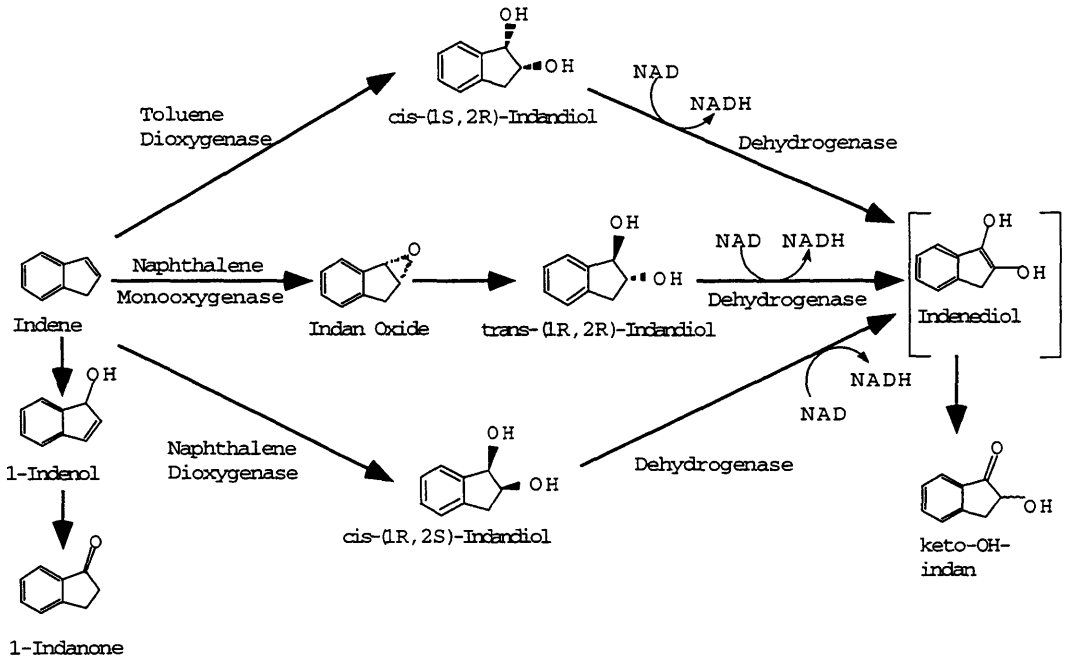


Figure 2-1: Proposed Indene Bioconversion Pathway in *Rhodococcus* I24
This pathway is adapted from Buckland et al. (1999) and Chatrain et al. (1998). The pathway was developed based on bioconversion analysis and induction studies. All products are detectable by HPLC except for indenediol. Indenediol is a proposed intermediate of the dehydrogenation reaction that isomerizes to keto-OH-indan.

Initial analysis of the system suggests that there are three reactions, possibly carried out by three different enzyme systems (or complexes), by which indene can be converted to a 1,2-indandiol. The presence in a given organism of multiple oxygenase enzymes capable of utilizing a common substrate is not uncommon (Asturias and Timmis 1993; Kosono et al. 1997). Chartrain and colleagues (Chartrain et al. 1998) have proposed that the bioconversion of indene in the I24 strain is conducted by a naphthalene-inducible dioxygenase, a toluene-inducible dioxygenase, and a naphthalene-inducible monooxygenase. An indene derived compound with the same retention time as chemically synthesized indan oxide is proposed to be produced in a monooxygenation reaction. The structural identity of this compound has not yet been determined. It is possible that the naphthalene-inducible monooxygenase activity is a second reaction of the proposed naphthalene dioxygenase since both reactions are inducible by naphthalene. Naphthalene dioxygenases have been characterized from *Pseudomonas* that can act both as dioxygenases and monooxygenases (Gibson et al. 1995). Interestingly, each of the proposed oxygenation reactions in *Rhodococcus* strain I24 seems to produce a specific indandiol stereoisomer suggesting discrete enzyme mechanisms. Other oxygenases have been found to act upon indene in a stereospecific manner (Gibson et al. 1995; Wackett et al. 1988).

In the proposed pathway shown in Figure 2-1, each of the oxygenation reactions is followed by a dehydrogenation reaction. These reactions may be carried out by a single dehydrogenase or by separate dehydrogenases whose expression and substrate specificity correlate with their respective dioxygenase or monooxygenase. We are interested in analyzing at the genetic level the oxygenation and dehydrogenation steps of this indene bioconversion in order to better understand and control the production of indandiol in this strain.

This chapter describes the initial genetic analyses of the *Rhodococcus* I24 genome and the *Rhodococcus* I24 indene bioconversion pathway. Using a functional screen for dioxygenase activity, we identify genes encoding the large and small subunit of a naphthalene-type dioxygenase, a dehydrogenase, and a putative aldolase. We demonstrate that the dioxygenase is capable of oxygenating indene. The implications of the findings on the proposed model for indene bioconversion in this strain and the relationship of these genes to previously identified dioxygenase genes are discussed.

Materials and Methods

Reagents

All strains and plasmids used in this study are listed in Table 2-1. All chemicals were reagent grade and purchased from Sigma (St. Louis, MO) or Aldrich Chemical Co (Milwaukee, WI) unless otherwise noted. All media components were purchased from Difco (Detroit, MI).

Table 2-1: Strains and Plasmids

Strains	*Description*	*Reference*
Rhodococcus I24	Converts indene to indandiol, orange colonies	Buckland et al, 1998
Rhodococcus SQ1	Easily transformable isolate of *R. erythroplis*	Quan and Dabbs, 1993
Escherichia coli JM109	*endA*1, *recA*1, *gyrA*96, *thi*, *hsdR*17, (r_k-, m_k+), *relA*1, *supE*44, λ-, *Δ(lac-proAB)*, [F', *traID36*, *proAB*, *lacI*^{q}ZΔM15]	Yanisch-Peron, C., Viera, J., and Messing, J., 1985
Plasmids		
SuperCOS	Amp^R, Kan^R, colE1 ori, COS sites	Stratagene
pEP2	Kan^R, NG2 ori	Zhang, 1994
pRhodoCOS	Kan^R, NG2 ori, COS sites	This study
pR4	Kan^R, NG2 ori, I24 genomic cosmid clone	This study
pR4-10	Kan^R, NG2 ori, *Sau*3a pR4 subclone	This study

Pulse Field Gel (PFG) Analysis

Pulse field gel analysis was carried out using a protocol modified from Lai and Birren (Lai and Birren 1990). Genomic DNA from *Rhodococcus* I24 was prepared for PFG analysis from cultures grown in LB medium to an OD_{600} of 0.8 –1.0 and treated with 0.2 mg/ml chloramphenicol and 0.01% isoniazid for two

hours. The cells and DNA were then treated as described (Lai and Birren 1990) except that the agitation step during 24-hour lysis was omitted. InCert® agarose (FMC, Philadelphia, PA) was used to embed the cells. To prepare for enzyme digestion, gel slices were equilibrated in TEN Buffer (10 mM Tris pH 8.0, 1 mM EDTA, 50 mM NaCl) overnight at room temperature. Individual gel slices were then treated with 10 mg of Pefabloc SC (Boehringer Mannheim, Indianapolis, IN) in 1 ml of TE pH 8.0 for 2 hours at 37°C. Gel slices were subsequently equilibrated in 45 ml of TEN Buffer for one hour to remove excess protease inhibitor. Gel slices were digested with *Ase*I and *Ssp*I (New England Biolabs, Beverly, MA) in their appropriate buffers and were incubated at 4°C overnight followed by a 6-8 hour incubation at 37°C.

The digested samples were analyzed in 1% agarose (Gibco BRL, Grand Island, NY) 0.5X TBE gels run in a BioRad Chef-DR II Pulse Field Gel Apparatus at 6 V/cm at 14°C for 16-19 hours at a ramp time of 15-75 seconds for *Ase*I and 1-40 seconds for *Ssp*I. Different ramp times were used to increase resolution over certain size ranges. For 20-350 kb fragments a ramp time of 1-25 seconds was used. A ramp time of 1-40 seconds resolved 20-450 kb fragments. Fragments of 250–850 kb were resolved at 15-75 seconds and 450-750 kb fragments by 30-60 seconds ramp time. The Lambda DNA PFG Ladder used as a molecular weight marker was obtained from New England Biolabs (Beverly, MA).

DNA Manipulation and Plasmid Construction

Restriction enzymes were purchased from New England Biolabs (Beverly, MA) and used according to the manufacturer's recommendations. Plasmid DNA was prepared using the Wizard Maxiprep Kit and the Wizard Miniprep Kit from Promega (Madison, WI), or boiling lysis miniprep (Sambrook et al. 1989).

To construct the cosmid vector pRhodoCOS, pEP2 (Zhang et al. 1994) was partially digested with *Hin*CII, then digested with *Bam*HI and treated with Shrimp Alkaline Phosphatase (Boehringer Mannheim, Indianapolis, IN). The 3.1 kb *Hin*CII-*Bam*HI fragment of pEP2 was ligated to the 4.2 kb *Bgl*II-*Sca*I fragment of SuperCOS (Stratagene, La Jolla, CA) to create pRhodoCOS. *E. coli* JM109 were transformed by electroporation (Sambrook et al. 1989) and plated on LB plates containing 100 µg/ml kanamycin. The pR4 subclones were made by a *Sau3A*I partial digest of 1.5 µg of pR4 DNA using 0.4 units of enzyme for 1 minute at 37°C followed by heat inactivation at 65°C for 20 minutes, followed by re-ligation. Subclones were analyzed be restriction enzyme digest with the following enzymes *Pst*I, *EcoR*I, *Not*I, *Cla*I, *Pvu*II, and *Nco*I.

Preparation of *Rhodococcus* Genomic DNA and Library

Genomic DNA was prepared from *Rhodococcus* as follows. A 200 ml LB culture was grown to saturation at 30°C. Cells were harvested by centrifugation for 10 min at 4000 x g; the supernatant was removed and the pellet was frozen at -20°C for 30-60 minutes. The pellet was resuspended in 5 ml TE with 20 mg lysozyme and 200 µg mutanolysin, and incubated for one hour at 37°C with shaking. One ml

of 0.5 M EDTA, 1 ml of 10% SDS, and 1 ml of 5M NaCl were added, and the cell suspension was mixed gently and incubated on ice for 10 min. The cell suspension was then treated with 2 mg of proteinase K for one hour at 37°C. Proteinase K treatment was followed by the addition of 3.77 g sodium perchlorate to the cell suspension which was mixed gently and incubated at room temperature with gentle agitation for 30 minutes. The cell suspension was extracted with one volume of phenol:chloroform:isoamyl alcohol (25:24:1), vortexed and centrifuged for 20 min at 4000 x g . The aqueous phase was extracted with 0.5 volumes of chloroform:isoamyl alcohol (24:1), vortexed and recentrifuged. To the aqueous phase two volumes ice cold 100% ethanol was added. The DNA was spooled onto a glass rod and resuspended in 5 ml TE. Next 50 µl RNase (Boehringer Mannheim, Indianapolis) was added and the sample was incubated at 37°C for 30 min. The solution was extracted with one volume of phenol:chloroform:isoamyl alcohol (25:24:1), vortexed and centrifuged for 20 minutes at 4000 x g. The aqueous phase was extracted with 0.5 volumes of chloroform:isoamyl alcohol (24:1), vortexed and recentrifuged. The aqueous phase was ethanol precipitated with two volumes of cold ethanol and 0.1 vol. of 3M sodium acetate, and the DNA was spooled on a glass rod and washed in 70% ethanol for 3 minutes. The DNA was air dried for 5-10 minutes and resuspended in TE. The I24 genomic DNA was quantitated and subjected to partial digest with 0.4 units of *Sau3A*I for 1 minute at 37°C and the restriction enzyme was heat inactivated at 65°C for 20 minutes.

The cosmid library of *Rhodococcus* I24 was made using pRhodoCOS following the Stratagene (La Jolla, CA) protocol provided with the SuperCOS vector kit and Gigapack III XL packaging extract. JM109 *E. coli* was used as the host strain for the cosmids (Yanisch-Perron et al. 1985). The resulting library contains 286 cosmid clones.

Screen for Dioxygenase Activity *In Vivo*

The cosmid clones in JM109 *E. coli* cells were grown on LB plates containing 100 µg/ml kanamycin at room temperature for three to four days and screened for blue color development. In *R. erythropolis* SQ1, blue color development was assayed by first growing colonies at 30°C. After the *Rhodococcus* colonies reached 2-3 mm in diameter, filter paper containing 600 µl 3% indole (w/v in DMF) was placed in the lid of petri dish and plate was sealed with parafilm. Plates were kept at room temperature for 24-48 hours to allow for blue color formation. The dimethylformamide (DMF) was purchased from Fisher Scientific (Fair Lawn, NJ).

Transformation of *R. erythropolis* SQ1

Competent *Rhodococcus* cells were prepared as follows. A 100 ml culture of *R. erythropolis* SQ1 in MB (5g/L yeast extract, 15 g/L Bacto-Tryptone, 5 g/L Bacto-soytone, 5 g/L NaCl), 1.5% glycine, 1.8% sucrose, 0.01% isoniazid was grown at 30°C to an OD_{600} of ~1.6-1.8 (Hewlett Packard Diode Array

Spectrophotometer 8452A). The culture was treated 1 µl of 100 mg/ml ampicillin and incubated at 30°C for one hour. The cells were harvested by centrifugation at 4000 x g for 10 minutes and washed twice with 30 ml cold EPB1 (20 mM Hepes pH 7.2, 5% glycerol). The washed cells were resuspended in 2 ml cold EPB2 (5 mM Hepes pH 7.2, 15% glycerol). Competent cells were stored at -80°C.

Electroporation using a BioRad Gene Pulser set to 2.50 kV, 400 Ohms, 25 Fd was used to transform 70 µl competent cells with ~1 µg DNA. Immediately after electroporation the cells were transferred to a microfuge tube and 400 µl of Recovery Broth (80 g/L Brain Heart Infusion mixed with equal volume of Solution 2 (80 g/L Sorbitol, 20 g/L Sucrose)) was added and incubated at 30°C for one hour without agitation. The transformation was plated on LB plates supplemented with 200 µg/ml kanamycin and grown at 30°C for 3-5 days.

Plasmid Miniprep from *Rhodococcus*

This protocol is based on a protocol from Vogt Singer and Finnerty (Vogt Singer and Finnerty 1988) with many modifications. LB cultures of *Rhodococcus* were grown to saturation at 30°C (approximately two days). A 1.5 ml sample was pelleted in a microfuge tube and the supernatant was discarded. The pellet was resuspended in 400 µl TENS (50 mM Tris HCl pH8.0, 10 mM EDTA, 50 mM NaCl, 20% sucrose) with 5 mg/ml fresh lysozyme and incubated at 37°C for 30 minutes. Then 185 µl 10%SDS and 30 µl of 1M Tris pH12.6 (adjusted with NaOH) was added. The suspension was incubated at 55°C for 30 minutes. 300 µl of Potassium acetate (3M potassium, 5 m acetate: 29.44 g potassium acetate, water to 88.5 ml, 11.5 ml glacial acetic acid) was added and the solution was incubated on ice for 5 minutes. The mixture was centrifuged for 20 minutes and the supernatant was transferred to a new tube. The DNA was precipitated with 0.6 volumes (450 µl) of isopropanol and centrifuged for 20 minutes. The pellet was washed in 70% ethanol and resuspended in 30 µl TE (10 mM Tris-HCl pH8.0, 1 mM EDTA).

Indene Bioconversion

Cultures containing 25 ml of LB, 5 ml of silicon oil, 150 µl indene, and 150 µg/ml kanamycin (if appropriate) were grown at 30°C for 4 days for end point bioconversion data. 1.0 ml samples of media were extracted with 6 ml of HPLC grade isopropanol (Mallinckrodt, Paris, KY) and 3 ml of milliQ water. Samples were vortexed and cleared of cells in a 5 min centrifugation at 4000 x g. The samples were then filtered through a 0.22 µm PVDF syringe filter (Alltech, Deerfield, IL) and 20 µl of the filtrate was injected onto a Zorbax Rx-C8 4.6x250 mm HPLC column on a Rainin Dynamax HPLC System. The HPLC protocol was followed as previously described (Chartrain et al. 1998). The indene, the silicon oil, and the HPLC standards *cis*-indandiol, *trans*-indandiol, 1-indenol, 1-indanone, and keto-hydroxy-indan were gifts from M. Chartrain (Merck Research Laboratories, Rahway, NJ). HPLC grade acetonitrile was purchased from Mallinckrodt (Paris, KY). Chirality of the *cis*-indandiol was determined by Chiral HPLC analysis as previously described (Chartrain et al. 1998).

DNA Sequencing and Analysis

Sequencing of pR4 was performed at the MIT Biopolymers Facility using an ABI cycle sequencer by primer walking. Primers were synthesized by Gibco/BRL Life Technologies (Grand Island, NY). pR4-10 was sequenced by primer walking by Lark Technologies Inc. (Houston, TX) using Big Dye™ Terminator Cycle Sequencing Reactions (PE-ABD). Reactions were analyzed on a 5.75% Long Ranger™ gels by an ABI 373A-S or on 5.0% Long Ranger™ gels by an ABI 377 automated sequencer. Approximately 43% of the sequence data are from both strands, 31% of the data are from multiple fragments in the same direction, and 26% is from single fragments.

All sequence data were analyzed using GenBank, EMBL, Swissprot databases, the BLASTN, BLASTP, and BLASTX (Altschul et al. 1990; Gish and State 1993) programs via the National Center for Biotechnology Information server, and PROSITE protein database (Bairoch 1992) via the ExPASy web site (http://expasy.hcuge.ch/sprot/prosite.html). Alignments, percent similarity, and phylogenetic analysis was carried using the Lasergene program (DNA star, Madison, WI) via a CLUSTAL alignment using a PAM250 residue weight table.

GenBank Accession Number

The GenBank number assigned to the DNA sequence data presented in this chapter is AF121905.

Results

Genome Size Analysis and Genomic Cosmid Library Construction

Rhodococcus strain I24 was previously isolated from soil contaminated with aromatic compounds (Buckland et al. 1999; Chartrain et al. 1998). Since this strain is not well characterized and nothing was known about the genomic content of the strain, we determined the approximate size of the genome using pulse-field gel analysis in order to facilitate further genetic analysis of the strain. This information is particularly pertinent to the construction of any genomic library. Given the fact that Rhodococci and related species are G+C rich (63-72%) (Williams et al. 1984), the enzymes *Ase*I (AT▼TAAT) and *Ssp*I (AAT▼ATT) were used in the PFG gel analysis because they are specific for AT rich regions and would likely digest the genomic DNA infrequently producing a reasonable number of fragments to quantitate. Figure 2-2 shows a representational pulse-field gel analysis of *Rhodococcus* I24 genomic DNA. By varying the ramp times we were able to increase the resolution over several size ranges to get a more accurate representation of the genomic content. The results from these analyses are shown in Table 2-2. Based on the data, the genome size is estimated to be approximately 3 Mb including a large plasmid of roughly 340 kb. This plasmid could be either linear or circular, and in single or multiple copies. This novel mega-plasmid has been named pI24.

Table 2-2: Pulse Field Gel Analysis of *Rhodococcus* I24

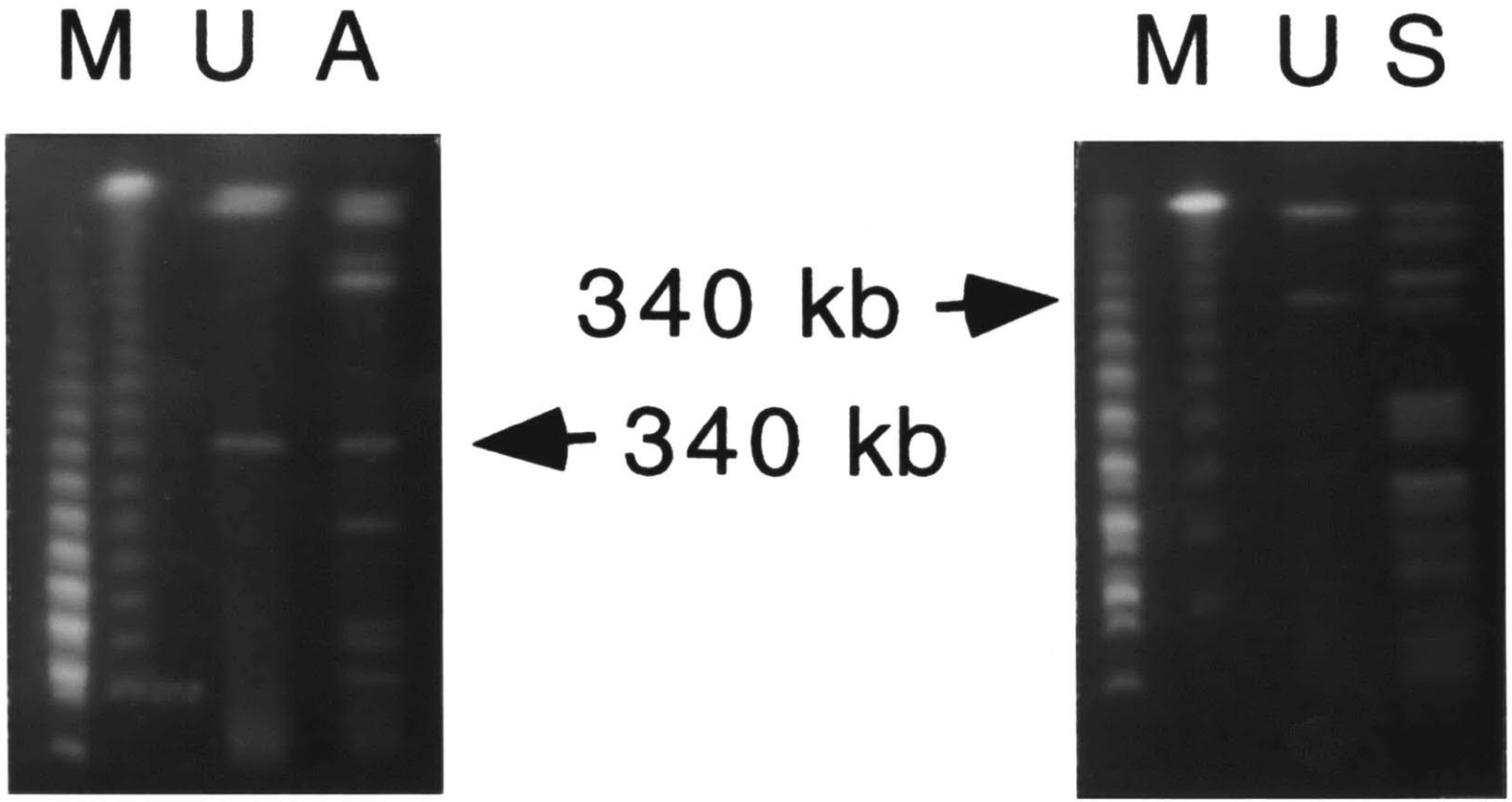
	AseI	*SspI*
Size of bands in kb	800	680
	380[a]	630[b]
	340	340
	210[a]	240
	195	130
	175	100
	130[a]	60
	110	
	95	
	90	
	65	
	60	
	35	
	25	
	20	
	15	
Total	2.75 Mb	2.8 Mb
Total with multiples	3.1-3.2 Mb	3.4[c] Mb

[a] Could be doublets due to intensity of ethidium bromide staining.
[b] At least a doublet. Used this in initial total.
[c] Total based on 630 kb band as a triplet

Figure 2-2: Pulsed-field Gel Analysis of *Rhodococcus* I24 (on next page)
The size of the I24 genome was analyzed by pulse-field gel analysis with two enzymes, *Ssp*I (S) and *Ase*I (A). Lambda PFG marker (M) is in lane 1 and Uncut (U) I24 genomic DNA is in lane 2. The I24 genome contains a large plasmid of 340 Kb.

Knowing the approximate size of the genome facilitated the construction of a genomic library. The first library attempted was an expression library. Construction of this library failed to produce clones with insert sizes greater than 1.0 kb. Since the screen (see below) was designed to find a functional unit comprised of more than one gene, this was unacceptable assuming that the average gene size is 1.0 kb and the operon in question likely was comprised four genes. One reason for this difficulty could be due to the high G+C content (~ 65% as determined by sequence analysis - see below) and the incompatibility of this G+C rich DNA in *Escherichia coli* whose G+C content is lower (~45-50%). Many different strain of *E. coli* were employed in the attempts to circumvent this problem; however, these experiments were unsuccessful.

To surmount this problem a genomic cosmid library was constructed. In order to prepare this library, a new cosmid vector was constructed called pRhodoCOS. pRhodoCOS contains two COS elements and an origin of replication that allows the plasmid to replicate in a broad range of hosts, including *E. coli* and *Rhodococcus*. A vector that could replicate in a broad host range was needed to account for the possibility that not all *Rhodococcus* genes would be expressed in *E. coli*. These genes would likely be expressed in another Gram-positive host. Others have shown that not all *Rhodococcus* genes are


M U A
M U S
340 kb
340 kb

expressed equally in *E. coli*, even from *E. coli* promoters (Asturias and Timmis 1993; Masai et al. 1995). Using this new vector a genomic library of *Rhodococcus* I24 was constructed containing 286 cosmid clones with approximately 30-40 kb inserts. Based on the genome size calculations, the library provides approximately three-fold coverage of the *Rhodococcus* I24 genome.

Identification of Ring-Hydroxylating Dioxygenase Activity

A functional screen was used to identify cosmid clones that were carrying genes encoding ring-hydroxylating dioxygenases. In some cases, other researchers have found that a ring-hydroxylating dioxygenase will convert indole to indoxyl, which spontaneously dimerizes to form indigo (Ensley et al. 1983; Hart et al. 1992). Figure 2-3 shows the indigo formation assay reaction. Indigo accumulation in the cells will cause the colonies to turn blue. *E. coli* is capable of converting tryptophan to indole via a tryptophanase, providing an internal source of indole. We screened 286 transformants of *E. coli* JM109, each carrying a single cosmid from the library, for blue color development. Nine of the 286 strains turned blue suggesting the presence of ring-hydroxylating dioxygenase activity. Cosmid DNA was isolated from these nine strains and analyzed with a panel of restriction enzymes. The results suggest that these nine clones contain overlapping portions of the genome (data not shown). The smallest clone, pR4, had an insert of 20-30 kb and was chosen for further study.

The indigo formation assay in *E. coli* provides a preliminary screen for ring-hydroxylating dioxygenase activity. This assay can be carried out in *Rhodococcus* provided that an external source of indole is provided and the host strain itself does not have any inherent detectable ring-hydroxylating dioxygenase activity (see Figure 2-3). A strain that does not change color in the presence of indole is *R. erythropolis* SQ1 (Kesseler et al. 1996; Quan and Dabbs 1993). To determine whether pR4 encoded ring-hydroxylating dioxygenase activity in *Rhodococcus*, the cosmid pR4 was transformed into *R. erythropolis* SQ1. We found that *R. erythropolis* SQ1 harboring the plasmid pR4 will turn blue when exposed to indole (Figure 2-4). This result strongly suggests that pR4 encodes a dioxygenase activity, and is consistent with the results obtained in *E. coli*.

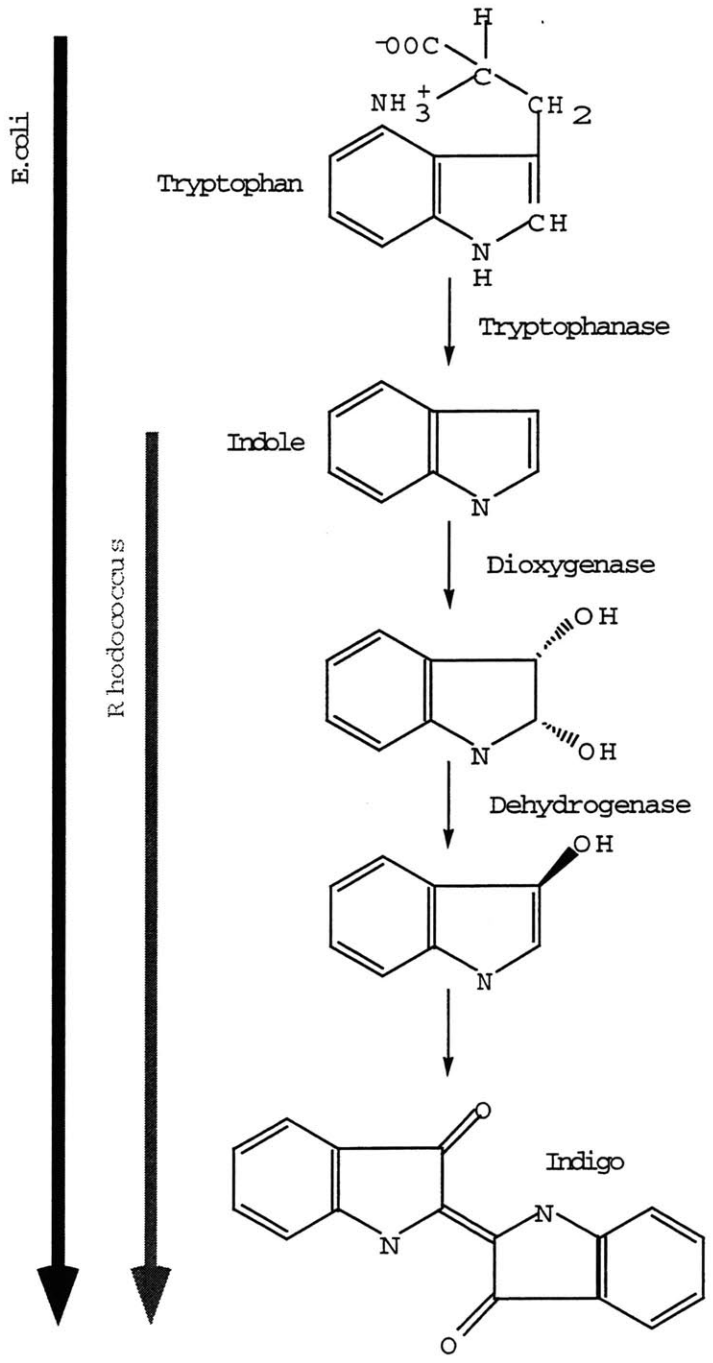


Figure 2-3: Indigo Formation Assay Reaction
Above is the reaction pathway from tryptophan to indigo for *E. coli*, and the pathway from indole to indigo for *Rhodococcus*. *E. coli* will take tryptophan from the media (e.g. LB media). *Rhodococcus* require the addition of indole in the vapor phase in order to carry out these reactions. Accumulated indigo in the cells (*E. coli* or *Rhodococcus*) will cause the cells to turn blue.

Figure 2-4: Indigo Formation Assay

Rhodococcus SQ1 cells alone, transformed with pR4, or transformed with pR4-10, are treated with 6% indole dissolved in dimethyl formamide (DMF) placed on filter paper in the lid of the plate. After a 24 hour incubation at room temperature the cells turn blue indicating indigo formation due to a possible ring-hydroxylating dioxygenase activity.

Indene Bioconversion Analysis

The primary screen in both *E. coli* and *R. erythropolis* SQ1 indicated that pR4 encoded an activity that permitted the conversion of indole to indigo. To further characterize this putative dioxygenase activity, we tested whether pR4 permitted oxygenation of indene by *R. erythropolis* SQ1. This is assayed in an indene bioconversion, a growth assay carried out in the presence of indene supplied in an oil phase. Indene is highly volatile and is not very miscible in water. The bioconversion of indene is measured by analysis of extracted media by high pressure liquid chromatography (HPLC). A sample chromatogram of the indene bioconversion products as visualized by HPLC is depicted in Figure 2-5.

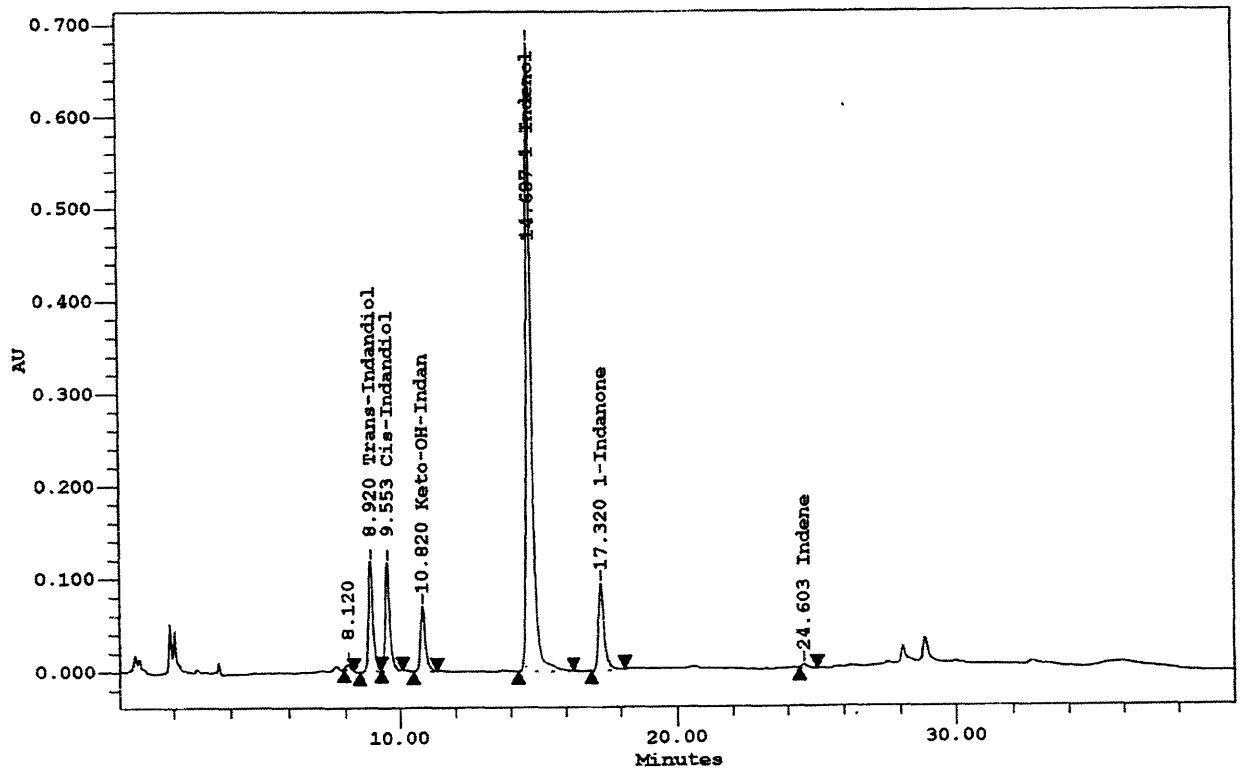


Figure 2-5: HPLC Indene Bioconversion Standards
This is a sample HPLC trace of the indene standards used in the analysis of indene bioconversion products. From left to right the compounds and retention times are: *trans*-indandiol (8.92), *cis*-indandiol (9.553), keto-OH-indan (10.82), indenol (14.587), indanone (17.32), and indene (24.603). Indan oxide, when available as a standard, would have a retention time of approximately 19.0 minutes. These are the approximate positions of the standards. Running conditions vary with each set of samples and column thereby shifting the retention times slightly.

In addition to being unable to convert indole to indigo, wildtype *R. erythropolis* SQ1 is unable to metabolize or oxygenate indene. Transformation of the pR4 cosmid into *R. erythropolis* SQ1 allows the strain to oxygenate indene (Figure 2-6). A series of bioconversion products were identified from *R. erythropolis* SQ1(pR4) including *cis*-indandiol, keto-hydroxy-indan, 1-indenol, and 1-indanone. The

production of *trans*-indandiol was not detected. Based on established pathways (Grund et al. 1995; Warhurst and Fewson 1994; Zylstra 1994; Zylstra et al. 1988) and the proposed pathway in *Rhodococcus* I24 (Buckland et al. 1999; Chartrain et al. 1998) (Figure 2-1), production of the *cis*-indandiol is indicative of dioxygenase activity. Keto-hydroxy-indan is produced by the dehydrogenation of the indandiol and its presence suggests that there is also a dehydrogenase activity on the pR4 cosmid. When compared to the product profile of *Rhodococcus* I24 (Figure 2-6), *R. erythropolis* SQ1(pR4) produces a subset of the products produced *Rhodococcus* I24 in an indene bioconversion. This suggests that pR4 encodes only some of the genes responsible for indene bioconversion in *Rhodococcus* I24.

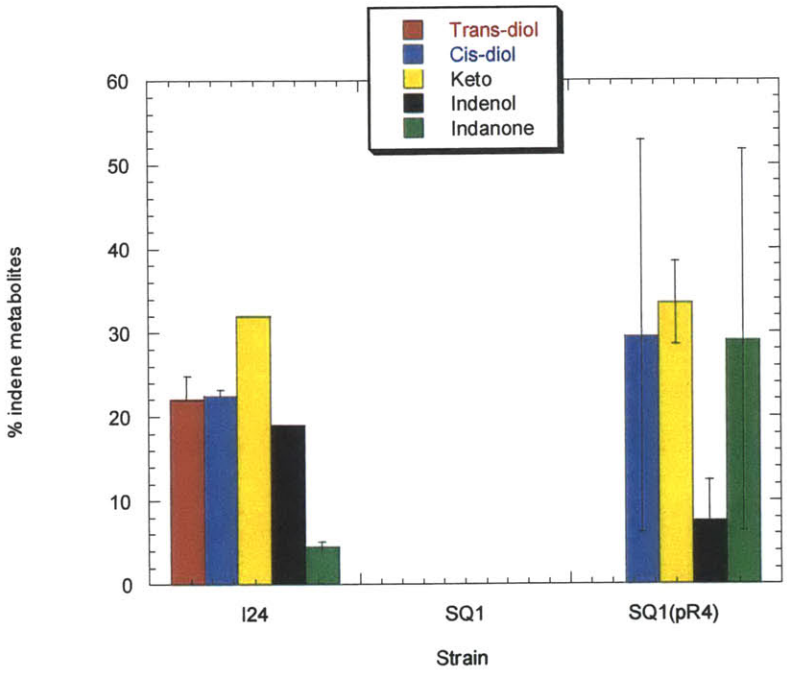


Figure 2-6: Indene Bioconversion Analysis of *Rhodococcus* SQ1(pR4)
The presence of indene metabolites were measured by HPLC analysis to determine if the cosmid clone pR4 contained a dioxygenase activity that would utilize indene as a substrate.

Rhodococcus I24 is known to produce two enantiomers of the *cis*-indandiol that are differentially induced by naphthalene and toluene (Chartrain et al. 1998). To further characterize the dioxygenase enzyme activity encoded by pR4, the chiral identity of the *cis*-indandiol produced by *R. erythropolis* SQ1(pR4) was examined by chiral HPLC analysis. This experiment was performed by M. Chartrain at Merck Research Laboratories (Rahway, NJ). The results of this analysis show that only *cis*-(1R,2S)-indandiol, not *cis*-(1S,2R)-indandiol, was produced (data not shown). Based on these results, we

conclude that pR4 carries a gene or genes that encode a specific dioxygenase activity and a possible dehydrogenase activity.

Sequence Analysis

The next step in this analysis was to identify which regions of pR4 contained the genes involved in the partial bioconversion of indene. Therefore, subclones of pR4 using a *Sau3A*I partial digest and re-ligation were constructed to reduce the size of the clone to something more manageable. The indigo formation assay was repeated to screen for subclones that retained ring-hydroxylating dioxygenase activity. Three clones were identified that retained indigo forming activity in the screen and were designated pR4-10, pR4-14, and pR4-20. Each subclone was subjected to an in depth restriction analysis to determine the approximate size of the insert. The insert size of pR4-20 was highly variable in this analysis therefore the clone was discarded. pR4-14 had an insert estimated to be 10.7 kb. Subclone pR4-10 appeared to have the smallest insert of approximately 7.0 kb in length and was chosen for sequence analysis. Furthermore subclone pR4-10, when transformed into *R. erythropolis* SQ1, promotes the conversion of indene to *cis*-indandiol (data not shown - see Chapter 3 for more analysis and discussion).

The entire insert of pR4-10 and approximately 4 kb of pR4 was subjected to DNA sequencing. The insert of subclone pR4-10 was found to be 7036 bp in length, confirming the restriction enzyme digest analysis. Consequently, approximately 11 kb of pR4 has been sequenced. The G+C content of the DNA sequenced was approximately 65%. The sequence data were also subjected to extensive BLAST searches using all of the available algorithms. Four open reading frames in the DNA sequenced were found that have significant homology to genes found in dioxygenase operons (Figure 2-7).

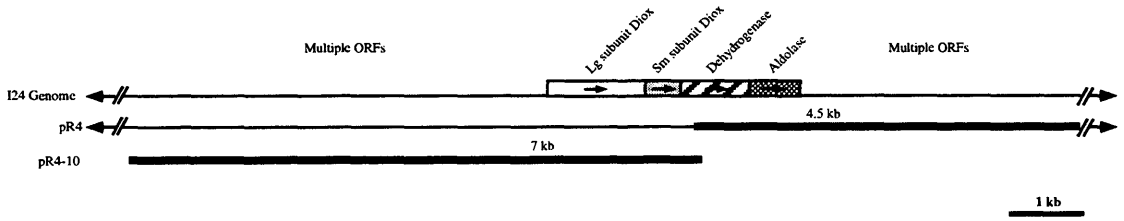


Figure 2-7: Sequence Analysis and Genome Structure
The structure of the region encoding the dioxygenase was determined by sequencing pR4 and pR4-10. The portion of pR4 sequenced is denoted by the thick line. In addition to the large and small subunits of the dioxygenase, the dehydrogenase and putative aldolase, multiple open reading frames (ORFs) were identified both upstream and downstream of the denoted gene cluster. None of these ORFs showed significant homology to known genes in BLAST searches of the GenEMBL databases.

The largest open reading frame detected is 1407 bp and shows homology to genes encoding the large subunits of a variety of dioxygenases. Based on the indene bioconversion results and the proposed

pathway, this gene is thought to correspond to the dioxygenase activity depicted in the lower portion of the pathway (Figure 2-1) that is naphthalene-inducible and produces *cis*-(1R,2S)-indandiol. Hence we have named this gene *nidA* for naphthalene-inducible dioxygenase system gene A. Analysis of the translated open reading frame shows a potential Rieske type iron-sulfur binding center that exactly matches the consensus sequence of C-x-H-R-G-x(8)-G-N-x(5)-C-x-Y-H-G found in proteins of this class (Figure 2-8) (Mason and Cammack 1992). Generally, the large subunits of these dioxygenases also have iron-binding sites comprised of two histidines and two tyrosines that are found toward the middle of the polypeptide; however, the spacing of these amino acids is not well conserved in the known bacterial dioxygenases. There are four histidines and three tyrosines near the middle of the polypeptide encoded by pR4, any of which could contribute to a iron-binding site.

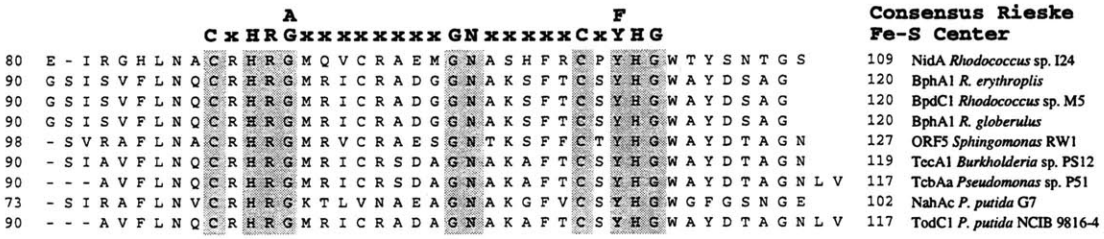


Figure 1-7: Rieske Iron-Sulfur Center of the NidA Dioxygenase Large Subunit
The consensus Rieske Iron-Sulfur Center is shown in bold at the top. Consensus residues are highlighted in grey. The NidA Dioxygenase Large Subunit has a region that matches the consensus (using PROSITE, Bairoch, 1992). Alignment was made using ClustalW with the Lasergene program (DNA star, Madison, WI). The GenBank accession numbers of the protein sequences are (from top to bottom): AF121905, D88020, U27591, X80041, AJ223219, U78099, U15298, 484406, J04996.

Next to *nidA* is a 519 bp open reading frame with homology to dioxygenase small subunit genes. We have named this open reading frame *nidB*. It has no distinguishing features or motifs. The small subunit of the dioxygenase is followed by another open reading frame of approximately 1.0 kb with homology to *cis*-dihydrodiol dehydrogenases and we have named it *nidC*. Analysis of the derived protein sequence shows that NidC has the characteristic C-terminal Y-x-x-x-K motif thought to be the active site of these dehydrogenases (Figure 2-8) (Jornvall et al. 1995; Krozowski 1994; Persson et al. 1991). The NidC dehydrogenase may also have a NAD^+ binding site in the N-terminal portion of the polypeptide. The characteristic NAD^+ binding site is G-x-x-x-G-x-G and there is a distinct G-x-G-x-G-x-G motif in the N-terminus of the dehydrogenase encoded by pR4 that aligns with other dehydrogenases (Figure 2-9)(Jornvall et al. 1995; Krozowski 1994; Persson et al. 1991).

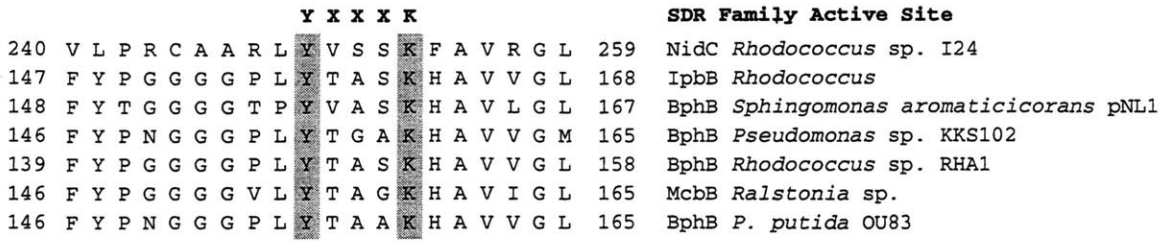


Figure 2-8: Dehydrogenase Active Site in NidC Dehydrogenase
The consensus active site is shown in bold at the top. Consensus residues are highlighted in grey.The active site is generally found in the middle towards the C-terminal portion of the protein. The NidC dehydrogenase has a region that matches the consensus (using PROSITE, Bairoch, 1992). This region lies in the C-terminal portion of the protein. Alignment was made using ClustalW with the Lasergene program (DNA star, Madison, WI). The GenBank accession numbers of the protein sequences are (from top to bottom): AF121905, AJ006127, AF07931, D17319, D32142, AJ006307, Y07655.

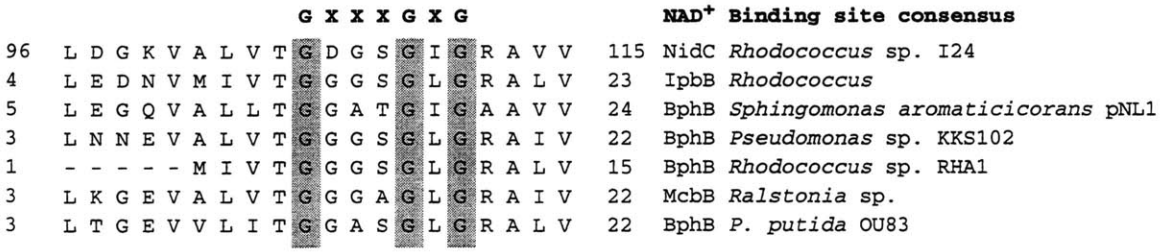


Figure 2-9: NAD+ Binding Site in NidC Dehydrogenase
The consensus binding site is shown in bold at the top. Consensus residues are highlighted in grey. The NidC dehydrogenase has a region that matches the consensus (using PROSITE, Bairoch, 1992). Alignment was made using ClustalW with the Lasergene program (DNA star, Madison, WI). The GenBank accession numbers of the protein sequences are (from top to bottom): AF121905, AJ006127, AF07931, D17319, D32142, AJ006307, Y07655.

The fourth open reading frame, which is located downstream of the *nidC* dehydrogenase gene, is 756 bp in length and exhibits weak homology to aldolase genes. We have named the putative aldolase gene *nidD*. Since we do not have an assay for this activity the name assignment is provisional. The name assignments of *nidABC* are based on homology and the indene bioconversion data. The regions upstream and downstream of these four genes (*nidABCD)* contain open reading frames that exhibit no significant homology to anything in the GenEMBL non-redundant databases. All four of the open reading frames identified by homology appear to have ribosome binding sites (RBS) associated with them based on the canonical RBS sites from *E. coli* (Shine and Dalgarno 1975).

All four of the gene identified by sequence homology were subjected to percent similarity and phylogenetic analysis of their DNA derived protein sequences. The derived protein sequences were

compared with the first five to seven most probable matches from the BLAST analyses. Figure 2-10 shows charts with the percent similarity of the Nid proteins to other similar proteins. On average, the Nid proteins are approximately 30% similar over the entire length of the protein to other aromatic degradation proteins. Notably, the average percent similarity for the putative aldolase is lower, around 24%. To get an idea of the evolutionary relationships of the Nid proteins to these similar proteins a phylogenetic analysis was conducted. It can be seen from Figure 2-11 that the Nid proteins diverge early from those proteins with which they are most closely related.

Figure 2-10: Sequence Similarity of NidABCD and Related Proteins

Sequence similarity analysis of aligned DNA derived protein sequences was determined. Alignment was performed using the CLUSTAL method with PAM250 residue weight table using MegAlign software from the Lasergene program (DNA star, Madison, WI). Sequence similarity is expressed as a percent and located to the right of the diagonal. A). The large subunit of the dioxygenase is analyzed. The GenBank accession numbers for the indicated genes are (from top to bottom): AF121905, D88020, U27591, X80041, AJ223219, U78099, U15298, 484406, J04996.. B). Analysis of the small subunit of the dioxygenase. GenBank accession numbers are (from top to bottom): AF121905, U24277, S51758, D32142, AJ223219, 2822266, Q52439, P08085, 3184045, 484406, J04996. C). The dehydrogenase is analyzed. GenBank accession numbers for the genes indicated are (from top to bottom): AJ006127, D32142, D17319, Y07655, AJ006307, AF07931, AF121905, AJ006126. D). Analysis of the aldolase. GenBank accession numbers are (from top to bottom): U09057, AB004059, AF010471, AF061751, AB000735, AF121905.

Figure 2-11: Phylogenetic Analysis of NidABCD and Related Proteins

Phylogenetic analysis of aligned DNA derived protein sequences with the highest homology to the *nid* genes was conducted. Alignment was performed using the CLUSTAL method with PAM250 residue weight table using the MegAlign software from the Lasergene program (DNA star, Madison, WI).. The x-axis of this unbalanced tree represents the number of substitution events, thus describing the divergence between sequences. A). The large subunit of the dioxygenase is analyzed. The GenBank accession numbers for the indicated genes are (from top to bottom): D88020, X80041, U27591, U78099, U15298, J04996, AJ223219, AF121905, 484406. B). Analysis of the small subunit of the dioxygenase. GenBank accession numbers are (from top to bottom): P08085, J04996, 3184045, U24277, D32142, S51758, 2822266, Q52439, AJ223219, AF121905, 484406. C). The dehydrogenase is analyzed. GenBank accession numbers for the genes indicated are (from top to bottom): AF121905, AJ006127, AF07931, D17319, D32142, AJ006307, AJ006126, Y07655. D). Analysis of the aldolase. GenBank accession numbers are (from top to bottom): AF121905, AB000735, AF061751, U09057, AB004059, AF010471.

Figure 2-10: Sequence similarity of NidABCD and related proteins

A. NidA Large Subunit of the Dioxygenase

Percent Similarity

	1	2	3	4	5	6	7	8	9		
1		29.2	29.2	29.2	28.7	31.0	29.9	26.9	29.3	1	NidA *Rhodococcus* sp. I24
2	126.9		99.4	99.8	34.2	72.1	71.6	28.5	71.6	2	BphA1 *R. erythroplis*
3	126.9	0.4		99.4	34.2	71.9	71.4	28.5	71.3	3	BpdC1 *Rhodococcus* sp. M5
4	126.9	0.0	0.4		34.2	72.1	71.6	28.5	71.6	4	BphA1 *R. globerulus*
5	142.1	99.5	99.5	99.5		33.4	33.5	30.3	34.4	5	ORF5 *Sphingomonas* RW1
6	126.3	33.7	34.0	33.7	104.1		95.3	29.4	86.7	6	TecA1 *Burkholderia* sp. PS12
7	128.0	33.2	33.5	33.2	102.9	3.7		29.8	88.7	7	TcbAa *Pseudomonas* sp. P51
8	160.9	136.2	137.3	136.2	139.6	138.0	139.7		29.6	8	NahAc *P. putida* G7
9	127.6	33.2	33.6	33.2	99.3	13.7	12.3	138.4		9	TodC1 *P. putida* NCIB 9816-4
	1	2	3	4	5	6	7	8	9		

B. NidB Small Subunit of the Dioxygenase

Percent Similarity

	1	2	3	4	5	6	7	8	9	10	11		
1		32.9	32.9	32.4	32.9	35.8	32.9	29.5	30.1	22.0	29.5	1	NidB *Rhodococcus* sp. I24
2	110.6		64.5	95.7	37.5	48.4	56.7	77.5	78.6	17.1	77.5	2	IpbA2 *R. erythroplis* BD2
3	110.6	45.8		66.1	38.1	52.2	56.5	63.4	64.0	18.3	63.4	3	BphA2 *R. globerulus*
4	112.8	4.4	43.0		38.6	50.0	56.7	79.1	79.7	16.6	79.1	4	BphA2 *Rhodococcus* sp. RHA1
5	102.8	95.0	96.0	93.2		39.2	43.2	40.3	39.8	21.6	40.3	5	ORF G6 *Sphingomonas* str.RW1
6	102.1	82.8	74.9	78.4	96.0		66.1	50.0	50.0	17.2	50.0	6	IpbAb *P. putida* RE204
7	112.8	62.4	61.8	62.4	84.6	43.0		56.7	55.6	21.2	56.7	7	BphA2 *Pseudomonas* sp. KKS102
8	127.6	26.7	46.8	24.5	93.2	78.4	62.4		93.0	17.6	100.0	8	BnzB *P. putida*
9	122.4	25.2	45.8	23.8	93.2	78.4	64.8	7.3		17.6	93.0	9	McbAb *Ralstonia* sp. JS705
10	160.3	212.0	200.0	212.0	197.0	200.0	183.2	202.0	202.0		17.6	10	NahAd *P. putida* G7
11	127.6	26.7	46.8	24.5	93.2	78.4	62.4	0.0	7.3	202.0		11	TodC2 *P. putida* NCIB 9816-4
	1	2	3	4	5	6	7	8	9	10	11		

C. NidC Dehydrogenase

Percent Similarity

	1	2	3	4	5	6	7	8		
1		33.7	37.2	33.3	33.8	42.8	12.3	31.0	1	NidC *Rhodococcus* sp. I24
2	115.0		45.9	54.8	98.1	55.6	11.5	55.2	2	IpbB *Rhodococcus*
3	106.1	84.1		41.0	46.4	54.5	10.9	41.4	3	BphB *Sphingomonas aromaticicorans* pNL1
4	116.0	66.5	97.8		54.0	67.9	11.6	72.5	4	BphB *Pseudomonas* sp. KKS102
5	117.1	0.4	84.7	66.6		54.5	12.2	54.8	5	BphB *Rhodococcus* sp. RHA1
6	94.5	64.8	67.2	41.8	63.7		15.5	62.6	6	McbB *Ralstonia* sp.
7	342.0	393.0	328.0	335.0	383.0	328.0		11.9	7	EdoC *Rhodococcus* sp.
8	123.4	65.7	99.6	33.8	65.8	51.5	309.0		8	BphB *P. putida* OU83
	1	2	3	4	5	6	7	8		

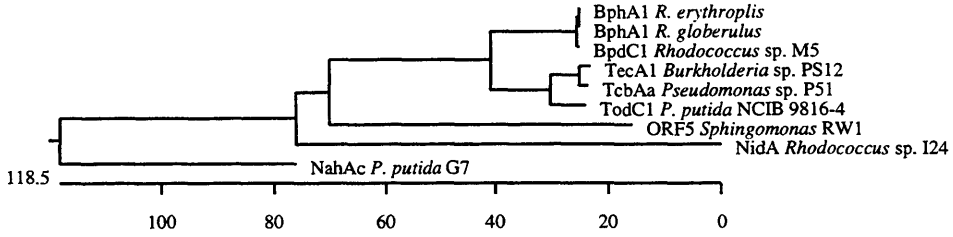
D. NidD Aldolase

Percent Similarity

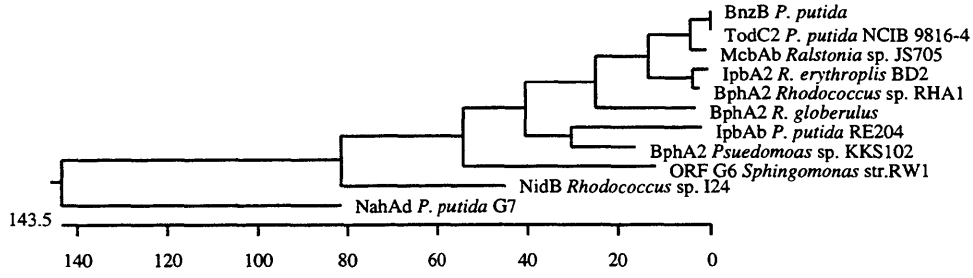
	1	2	3	4	5	6		
1		26.2	24.2	24.2	23.8	23.8	1	NidD *Rhodococcus* sp. I24
2	145.2		35.9	36.5	37.1	36.5	2	PhdJ *Nocardioides* sp. KP7
3	141.6	111.9		74.2	73.9	73.3	3	PhnE *Burkholderia* sp. RP007
4	158.9	111.2	30.4		98.2	95.8	4	NahE *Pseudomonas putida* plasmid
5	160.5	109.9	31.1	1.8		95.2	5	PahE *P. putida* st. OUS82
6	160.5	113.4	31.6	4.4	5.0		6	NahD *P. putida*
	1	2	3	4	5	6		

Figure 2-11: Phylogenetic analysis of NidABCD and related proteins

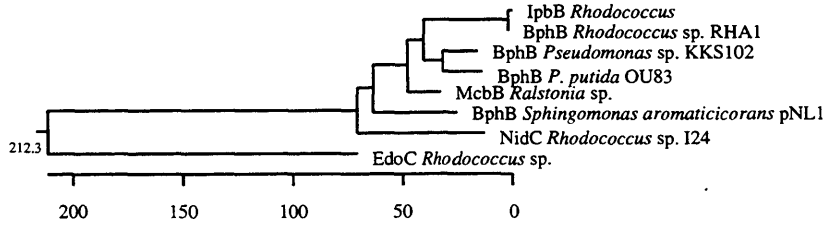
A. Large Subunit of the Dioxygenase



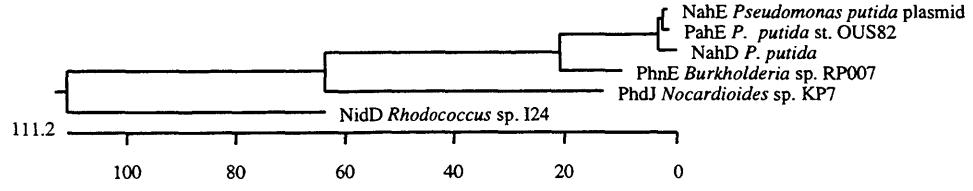
B. Small Subunit of the Dioxygenase



C. Diol Dehydrogenase



D. Aldolase



Discussion

Genomic analysis of *Rhodococcus* I24 by pulsed-field gel electrophoresis suggested that the genome was approximately 3 Mb. Other strains of *Rhodococcus* and related organisms such as *Corynebacterium* and *Mycobacterium* have been analyzed in this fashion. Genome sizes of actinomycetes appear to range from 3 Mb to 7 Mb [Bigey, 1995 #521; (Pisabarro et al. 1998); Correia, 1994 #519; Bathe, 1996 #520; Philipp, 1996 #522]. The size of *Rhodococcus* I24 is within the range of those strains previously investigated, although it is smaller than other *Rhodococcus* strains analyzed to date. A large plasmid named pI24 was identified in *Rhodococcus* I24. It is unknown whether this plasmid is linear or circular. Other have shown that Gram-positive and Gram-negative bacteria can have linear plasmids [Dabrock, 1994 #508; (Pisabarro et al. 1998); Picardeau, 1997 #498; Hinnebusch, 1993 #523] and circular plasmids (Dabbs 1988) as part of their genomic content. To determine the structure of the pI24 plasmid further investigation is needed.

Using a functional assay in *E. coli* and DNA sequencing an 11.0 kb fragment of *Rhodococcus* I24 DNA was identified that encoded four genes (*nidABCD)* including a dioxygenase and a dehydrogenase capable of breaking down indene to *cis*-(1R,2S)-indandiol, 1-indenol, 1-indanone, and keto-hydroxy-indan. Nine overlapping cosmid clones were identified that carry these genes. We believe that these genes could be located on the large plasmid found in the strain. This plasmid could be in multiple copies thereby accounting for the disproportionate number of clones identified. A preliminary investigation of this possibility was carried out via Southern blotting of pulse-field gel separated DNA using the dioxygenase gene *nidA* as a probe. Other aromatic compound degrading bacteria, including *Pseudomonas*, *Burkholderia*, and *Rhodococcus*, have been found to have operons for aromatic catabolism located on a plasmid separate from the rest of the genome (Aemprapa and Williams 1998; Dabrock et al. 1994; Fong et al. 1996; Masai et al. 1997; Sanseverino et al. 1993; Shields et al. 1995). The preliminary results suggested that the *nidA* dioxygenase may be located on the plasmid (data not shown). However, due to a variety of technical issues this result has not been confirmed. Future experiments have been planned to address this issue including plasmid isolation and screening for a plasmid-cured strain. It may also be possible to transfer pI24 to another strain by conjugation. Often these large plasmids containing aromatic degradation pathways can be transferred from one strain to another by transconjugation (Nakazawa and Yokota 1977; Shields et al. 1995; Williams and Worsey 1976). This method of analysis is being explored by trying to transfer pI24 to the heterologous host *R. erythropolis* SQ1 and screening for indene bioconversion or indigo formation.

Bioconversion analysis suggests that the genes *nidAB* correspond to the subunits of the naphthalene-inducible dioxygenase proposed for *Rhodococcus* I24 (Chartrain et al. 1998). When the plasmid pR4 is transformed into the naïve host *R. erythropolis* SQ1, the transformed strain only produces the *cis*-(1R,2S)-indandiol when provided indene as a substrate (Figure 1). This finding strengthens the

hypothesis that multiple dioxygenases produce the different indandiols in *Rhodococcus* I24 in a stereospecific manner (Buckland et al. 1999; Chartrain et al. 1998). The lack of *cis*-(1S,2R)-indandiol and *trans*-(1R,2R)-indandiol production by the naïve *R. erythropolis* SQ1 containing pR4 also supports this notion. The absence of *cis*-(1S,2R)-indandiol suggests that there is a separate oxygenase that produces this enantiomer. We believe this to be a toluene-inducible dioxygenase (Chartrain et al. 1998). The fact that no *trans*-(1R,2R)-indandiol was produced by *R. erythropolis* SQ1(pR4) suggests that there is a third enzyme in *Rhodococcus* I24 carrying out an oxygenation reaction producing this compound. This third enzyme could be a monooxygenase that is inducible by naphthalene (Chartrain et al. 1998). Other evidence that is consistent with our proposal that pR4 encodes a functional dioxygenase is the production of indenol and indanone. It has previously been shown that indenol and indanone can be produced by both naphthalene and toluene dioxygenases using indene as a substrate (Gibson et al. 1995; Wackett et al. 1988). The formation of indenol and indanone in *Rhodococcus* I24 can now be attributed, at least partially, to the *nidAB* dioxygenase we have identified. It is possible that the other oxygenases in the system can contribute to the formation of these products as well. We are currently investigating substrate utilization, gene deletion and complementation studies with the *nidAB* dioxygenase genes so that we may control and modify this bioconversion pathway (see Chapter 4).

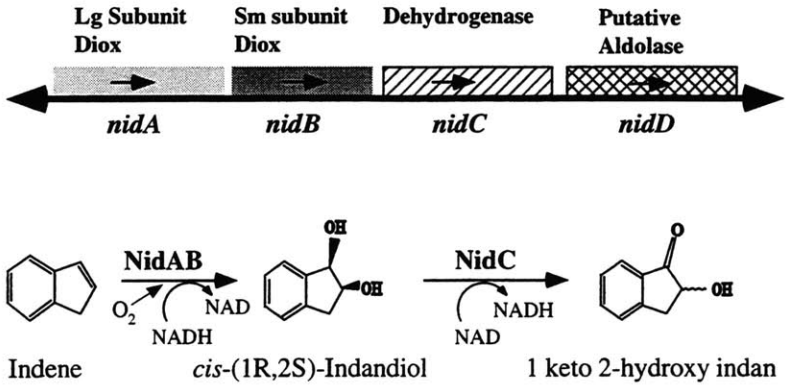


Figure 2-12: Genes and Compounds

Depicted above are the genes identified on the cosmid clone pR4 from *Rhodococcus* I24. The name *nid* stands for naphthalene inducible dioxygenase system. When these genes are expressed in the naïve host *R. erythroplis* SQ1 they enable the strain to carry out the reactions shown above. All of the compounds can be detected by HPLC.

Bioconversion analysis suggests that pR4 carries a gene encoding a diol dehydrogenase. This is evidenced by the fact that there is keto-OH-indan produced during a bioconversion using the strain *R. erythropolis* SQ1(pR4). Chemically, keto-OH-indan can be produced by the oxidation of *cis*-(1R,2S)-indandiol. Furthermore, the gene *nidC* identified by DNA sequencing of pR4 has homology to diol dehydrogenases. Overall, the indene bioconversion analysis in conjunction with DNA sequence analysis

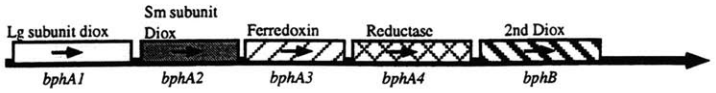
has led to the model shown in Figure 2-12 showing the genes, their relative positions and the chemical reactions for which the protein products are proposed to be responsible for.

Sequence analysis strongly suggests that the *nidABCD* genes belong to a family of genes capable of degrading aromatic compounds. Comparison of the *nidABCD* dioxygenase system with other dioxygenase systems in their overall gene organization suggests that the *nid* gene organization and genes in *Rhodococcus* I24 are different from other dioxygenase operons. Some examples of the gene organization of operons involved in aromatic compound degradation are given in Figure 2-13. By sequence homology, these operons are some of the most closely related to the *nidABCD* genes. A general pattern can be seen in the gene organizations. The *nid* genes appear to be divergent from this general pattern. When the homologous sequences are aligned and subjected to phylogenetic analysis the *nid* genes diverge early from their most homologous relatives (Figure 2-11). This is consistent with the divergence seen in the organization of the open reading frames.

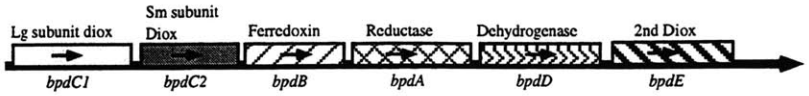
One way the *nid* genes are different from the general pattern is in the lack of an identifiable reductase, ferredoxin, or catechol dioxygenase within the 11 kb of DNA sequenced. The ferredoxin and the reductase components are required for electron transfer to the dioxygenase components in systems that are similar to *nidAB* (Mason and Cammack 1992). Therefore it is reasonable to suspect that these two components are needed for the dioxygenase system from pR4. This suggests that there may be as yet unidentified genes for the ferredoxin, the reductase on pR4 or that these components are unnecessary for this dioxygenase system. The region of pR4 potentially carrying these unidentified genes has been partially delineated by the pR4-10 subclone since this subclone retains indene bioconversion activity (Figure 2-7, and Chapter 3). Another possibility is that the *nidAB* dioxygenase subunits can borrow the ferredoxin and reductase components from the *R. erythropolis* SQ1 host strain. It has been previously demonstrated that the dioxygenase subunits from one system or hybrid dioxygenase subunits, can borrow the ferredoxin and reductase components of another system thereby altering the activity of the dioxygenase system (Beil et al. 1998; Furukawa et al. 1993; Hirose et al. 1994). In addition, it has been demonstrated in *Pseudomonas* that the reductase component can be located elsewhere in the operon, separated from the terminal dioxygenase subunits by a number of intervening genes (Kikuchi et al. 1994), or scattered in the genome (Armengaud et al. 1998). However, if the later were the case in *Rhodococcus* I24 a functional dioxygenase might not be identified in our functional screen. There are plans for others to further investigate the unidentified open reading frames in the I24 *nid* region in functional studies such as the creation of nested deletions to determine the minimal functional unit.

The inducible production of different stereoisomers of indandiol by *nidAB* and the other oxygenases implies that these enzymes may be subtly different from each other. Since indene cannot be used as a carbon source by the I24 strain (Chartrain et al. 1998) and is therefore not likely a natural substrate for these enzymes, the different enantiomers of indandiol may be the consequence of steric constraints on

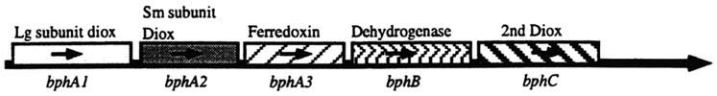
Rhodococcus globerulus biphenyl dioxygenase *(bph)* [a]



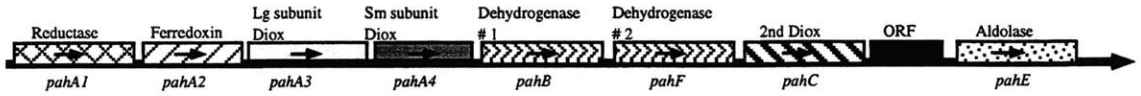
Rhodococcus sp. M5 biphenyl dioxygenase *(bpd)* [b]



Pseudomonas sp. biphenyl dioxygenase *(bph)* [c]



Pseudomonas aeroginosa naphthalene dioxygenase *(pah)* [d]



Rhodococcus I24 naphthalene-inducible dioxygenase *(nid)*

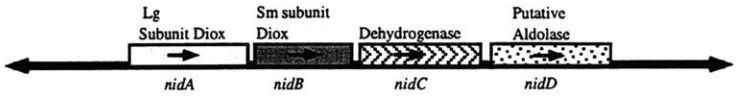


Figure 2-13: Comparison of Gene Organization of Dioxygenase Systems

Four different dioxygenase systems that are similar to the *Rhodococcus* I24 *nid region* system by homology are compared to each other by their gene organization.
[a] Asturias, J. A., E. Diaz, and K. N. Timmis. Gene 1995 156(1): 11-18
[b] Wang, Y. et al. Gene 1995 Oct 164(1): 117-122
[c] Fukuda, M., et al. Biochem Biophys Res Commun 1994 July 202(2): 850-856
[d] Takizawa, N. T. Iida, K. Yamauchi, S. Satoh, Y. Wang, M. Fukuda, and H. Kiyohara. Direct submission to Genbank, accession number: D84146

indene in the active site of these enzymes. Alternatively, the oxygenases could have different mechanisms of oxygenation resulting in the production of a specific enantiomer. This has been partially demonstrated for the toluene monooxygenases. The final product of the upper pathway of toluene degradation is the same in many strains, but the method by which toluene is oxygenated by a monooxygenase varies (Zylstra 1994). The enzyme mechanisms of previously identified bacterial dioxygenases have not yet been determined. We are in the process of identifying the genes for the other oxygenases in I24 in order to further investigate the stereospecificity of product formation.

Many strains of *Rhodococcus* are capable of breaking down aromatic compounds. *Rhodococcus* I24 is unique in that it employs at least three different types of oxygenases capable of acting upon indene.

This bioconversion is stereospecific and inducible. The gene structure for the *Rhodococcus* I24 naphthalene-inducible dioxygenase is unlike any dioxygenase gene cluster identified to date. We are currently investigating the other genes on this cosmid clone in order to better understand this dioxygenase system and indene bioconversion in *Rhodococcus* I24. In our study of bacteria capable of stereospecific bioconversion, we hope to develop efficient and controllable methods of producing important chiral substances and to gain a better understanding of how this organism and others like it metabolize aromatic substrates.

Acknowledgements

This work was supported by a grant from Merck Research Laboratories. S. L. T. was supported in part by NIH Biotechnology Training Grant # 2T32 GM08334-10.

We would like to thank Merck Research Laboratories for the strains, some of the sequencing, and chiral analysis. We also thank members of the Sinskey Laboratory for their discussion, comments and support. We especially thank P.A. Lessard for plasmids, K Yanagimachi for some HPLC analysis, and E. Lankenau for assistance with the PFG analysis. We also thank A. Chapman, M. Hillis, and I. Kim for all of their assistance.

References

Aemprapa S, Williams PA (1998) Implications of the *xylQ* gene of TOL plasmid pWW102 for the evolution of aromatic catabolic pathways. Microbiology 144: 1387-96

Allen CCR, Boyd DR, Larkin MJ, Reid KA, Sharma ND, Wilson K (1997) Metabolism of Napthalene, 1-Napthol, Indene, and Indole by *Rhodococcus* sp. Strain NCIMB 12038. Appl Environ Microbiol 63: 151-155

Altschul SF, Gish W, Miller W, Myers EW, Lipman AJ (1990) Basic Local Alignment Search Tool. J Mol Biol 215: 403-410

Armengaud J, Happe B, Timmis KN (1998) Genetic analysis of dioxin dioxygenase of *Sphingomonas* sp. Strain RW1: catabolic genes dispersed on the genome. J Bacteriol 180: 3954-66

Asturias JA, Timmis KN (1993) Three Different 2,3-Dihydroxybiphenyl-1,2-Dioxygenase Genes in the Gram-Positive Polychlorobiphenyl-Degrading Bacterium *Rhodococcus globerulus* P6. J Bacteriol 175: 4631-4640

Bairoch A (1992) PROSITE: A dictionary of sites and patterns in proteins. Nucleic Acids Research 20: 2013-2018

Beil S, Mason JR, Timmis KN, Pieper DH (1998) Identification of chlorobenzene dioxygenase sequence elements involved in dechlorination of 1,2,4,5-tetrachlorobenzene [In Process Citation]. J Bacteriol 180: 5520-8

Buckland B, Drew S, Connors N, Chartrain M, Lee C, Salmon P, Gbewonyo K, Gailliot P, Singhvi R, Olewinski R, Sun W, Reddy J, Zhang J, Zhou W, Jackey B, Goklen K, Junker B, Greasham R (1999)

Microbial conversion of Indene to Indandiol, a key intermediate in the synthesis of CRIXIVAN. Metabolic Engineering 1: 63-74

Carless HAJ (1992) The use of cyclohexa-3,5-diene-1,2-diols in enantiospecific synthesis. Tetrahedron: Asymmetry 3: 795-826

Chartrain M, Jackey B, Taylor C, Sandford V, Gbewonyo K, Lister L, DiMichele L, Hirsch C, Heimbuch B, Maxwell C, Pascoe D, Buckland B, Greasham R (1998) Bioconversion of indene to *cis*-(1S,2R)-indandiol and *trans*-(1R,2R)-indandiol by *Rhodococcus* species. J Ferment Bioeng 86: 550-558

Collins AN, Sheldrake GN, Crosby J (1997) Chirality in Industry II: Developments in Manufacturing and Applications of Optically Active Compounds. John Wiley & Sons, Inc., New York

Dabbs ER, Gillian J. Sole (1988) Plasmid-borne resistance to arsenate, arsenite, cadmium, and chloramphenicol in a *Rhodococcus* species. Mol Gen Genet 211: 148-154

Dabrock B, Kesseler M, Averhoff B, Gottschalk G (1994) Identification and characterization of a transmissible linear plasmid from *Rhodococcus erythropolis* BD2 that encodes isopropylbenzene and trichloroethene catabolism. Appl Environ Microbiol 60: 853-60

Ensley BD, Ratzkin BJ, Osslund TD, Simon MJ, Wackett LP, Gibson DT (1983) Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. Science 222: 167-9

Fong KP, Goh CB, Tan HM (1996) Characterization and expression of the plasmid-borne *bedD* gene from *Pseudomonas putida* ML2, which codes for a NAD+-dependent cis-benzene dihydrodiol dehydrogenase. J Bacteriol 178: 5592-601

Furukawa K, Hirose J, Suyama A, Zaiki T, Hayashida S (1993) Gene components responsible for discrete substrate specificity in the metabolism of biphenyl (*bph* operon) and toluene (*tod* operon). J Bacteriol 175: 5224-32

Gibson DT, Resnick SM, Lee K, Brand JM, Torok DS, Wackett LP, Schocken MJ, Haigler BE (1995) Desaturation, Dioxygenation, and Monooxygenation Reactions Catalyzed by Napthalene Dioxygenase from *Pseudomonas* sp. Strain 9816-4. J Bacteriol 177: 2615-2621

Gish W, State DJ (1993) Identification of protein coding regions by database similarity search. Nature Genet 3: 266-272

Grund E, Schmitz A, Fiedler J, Gartemann K-H (1995) Aerobic biodegradation of polycyclic and halogenated aromatic compounds. In: Singh VP (ed) Biotransformations: Microbial Degradation of Health Risk Compounds. Elsevier Science B.V, New York

Hart S, Koch KR, Woods DR (1992) Identification of indigo-related pigments produced by *Escherichia coli* containing a cloned *Rhodococcus* gene. J Gen Microbiol 138: 211-6

Hirose J, Suyama A, Hayashida S, Furukawa K (1994) Construction of hybrid biphenyl (*bph*) and toluene (*tod*) genes for functional analysis of aromatic ring dioxygenases. Gene 138: 27-33

Jornvall H, Persson B, Krook M, Atrian S, Gonzalez-Duarte R, Jeffery J, Ghosh D (1995) Short-chain dehydrogenases/reductases (SDR). Biochemistry 34: 6003-13

Kesseler M, Dabbs ER, Averhoff B, Gottschalk G (1996) Studies on the isopropylbenzene 2,3-dioxygenase and the 3-isopropylcatechol 2,3-dioxygenase genes encoded by the linear plasmid of *Rhodococcus erythropolis* BD2. Microbiology 142: 3241-3251

Kikuchi Y, Nagata Y, Hinata M, Kimbara K, Fukuda M, Yano K, Takagi M (1994) Identification of the bphA4 gene encoding ferredoxin reductase involved in biphenyl and polychlorinated biphenyl degradation in *Pseudomonas* sp. strain KKS102. J Bacteriol 176: 1689-94

Kosono S, Maeda M, Fuji F, Arai H, Kudo T (1997) Three of the Seven *bphC* Genes of *Rhodococcus erythropolis* TA421, Isolated from a Termite Ecosystem, Are Located on an Indigenous Plasmid Associated with Biphenyl Degradation. Appl Environ Microbiol 63: 3282-3285

Krozowski Z (1994) The short-chain alcohol dehydrogenase superfamily: variations on a common theme. J Steroid Biochem Mol Biol 51: 125-30

Lai E, Birren B (1990) Electrophoresis of large DNA molecules: theory and applications. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Masai E, Sugiyama K, Iwashita N, Shimizu S, Hauschild JE, Hatta T, Kimbara K, Yano K, Fukuda M (1997) The *bphDEF* meta-cleavage pathway genes involved in biphenyl/polychlorinated biphenyl degradation are located on a linear plasmid and separated from the initial *bphACB* genes in *Rhodococcus* sp. strain RHA1. Gene 187: 141-149

Masai E, Yamada A, Healy JM, Hatta T, Kimbara K, Fukuda M, Yano K (1995) Characterization of Biphenyl Catabolic Genes of Gram-Positive Polychlorinated Biphenyl Degrader *Rhodococcus* sp. Strain RHA1. Appl Environ Microbiol 61: 2079-2085

Mason JR, Cammack R (1992) The electron-transport proteins of hydroxylating bacterial dioxygenases. Annu Rev Microbiol 46: 277-305

Nakazawa T, Yokota T (1977) Isolation of a mutant TOL plasmid with increased activity and transmissibility from *Pseudomonas putida (arvilla)* mt-2. J Bacteriol 129: 39-46

Persson B, Krook M, Jornvall H (1991) Characteristics of short-chain alcohol dehydrogenases and related enzymes. Eur J Biochem 200: 537-43

Pisabarro A, Correia A, Martin JF (1998) Pulsed-field gel electrophoresis analysis of the genome of *Rhodococcus fascians*: genome size and linear and circular replicon composition in virulent and avirulent strains. Curr Microbiol 36: 302-8

Quan S, Dabbs ER (1993) Nocardioform arsenic resistance plasmid characterization and improved *Rhodococcus* cloning vectors. Plasmid 29: 74-9

Sambrook J, Fritsch E, Maniatis T (1989) Molecular Cloning: A Laboratory Manual, 2nd edition edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sanseverino J, Applegate BM, King JM, Sayler GS (1993) Plasmid-mediated mineralization of naphthalene, phenanthrene, and anthracene. Appl Environ Microbiol 59: 1931-7

Shields MS, Reagin MJ, Gerger RR, Campbell R, Somerville C (1995) TOM, a new aromatic degradative plasmid from *Burkholderia (Pseudomonas) cepacia* G4. Appl Environ Microbiol 61: 1352-6

Shine J, Dalgarno L (1975) Determinant of cistron specificity in bacterial ribosomes. Nature 254: 34-38

Vogt Singer ME, Finnerty WR (1988) Construction of an *Escherichia coli-Rhodococcus* Shuttle Vector and Plasmid Transformation in *Rhodococcus* spp. J Bacteriol 170: 638-645

Wackett LP, Kwart LD, Gibson DT (1988) Benzylic monooxygenation catalyzed by toluene dioxygenase from *Pseudomonas putida*. Biochemistry 27: 1360-7

Warhurst AM, Fewson CA (1994) Biotransformations catalyzed by the genus *Rhodococcus*. Crit Rev Biotechnol 14: 29-73

Williams PA, Worsey MJ (1976) Ubiquity of plasmids in coding for toluene and xylene metabolism in soil bacteria: evidence for the existence of new TOL plasmids. J Bacteriol 125: 818-28

Williams ST, Sharpe ME, Holt JG (1984) Bergey's Manual of Systematic Bacteriology. , vol 4. Williams and Wilkins, Baltimore

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-19

Zhang Y, Praszkier J, Hodgson A, Pittard AJ (1994) Molecular Analysis and Characterization of a Broad-Host-Range Plasmid, pEP2. J Bacteriol 176: 5718-5728

Zylstra GJ (1994) Molecular Analysis of Aromatic Hydrocarbon Degradation. In: Garte SJ (ed) Molecular Environmental Biology. Lewis Publishers, Boca Raton, pp 83-115

Zylstra GJ, McCombie WR, Gibson DT, Finette BA (1988) Toluene degradation by *Pseudomonas putida* F1: genetic organization of the tod operon. Appl Environ Microbiol 54: 1498-503

Zylstra GJ, Wackett LP, Gibson DT (1989) Trichloroethylene degradation by *Escherichia coli* containing the cloned *Pseudomonas putida* F1 toluene dioxygenase genes. Appl Environ Microbiol 55: 3162-6

Chapter Three: Characterization of the *nidC* Diol Dehydrogenase from *Rhodococcus* strain I24

Abstract

Part of the indene bioconversion pathway in *Rhodococcus* strain I24 is carried out by the protein products encoded by the genes *nidABC*. Genetic and biochemical analyses demonstrate that the *nidC* gene encodes a functional diol dehydrogenase. The NidC dehydrogenase is stereospecific for *cis*-(1R,2S)-indandiol, one of three indandiols produced during an indene bioconversion in *Rhodococcus* strain I24. In *nidC* deletion mutant studies we uncovered a second diol dehydrogenase activity capable of utilizing *cis*-(1R,2S)-indandiol as a substrate. This underlying activity is naphthalene-inducible.

Introduction

Bacteria, with their highly evolved ability to metabolize a wide variety of compounds, are a resource for enantiospecific precursors for biologically active products (Collins 1997; Patel 1997). When indene is supplied as a substrate, *Rhodococcus* strain I24 is capable of producing two compounds which can be used as precursors to the HIV protease inhibitor indinavir sulfate (Buckland et al. 1999; Chartrain et al. 1998). Figure 3-1 shows the proposed pathway of indene bioconversion in this system. Chartrain and colleagues proposed that *Rhodococcus* strain I24 produces three different indandiols via three separate stereospecific oxygenase systems (Chartrain et al. 1998). Many dioxygenase systems have been shown to exhibit stereospecific oxygenation (Boyd et al. 1989; Hamberg et al. 1994; Lee 1997; Resnick 1994). Some bacteria, including *Pseudomonas* and *Rhodococcus*, have dioxygenase systems that exhibit stereospecific oxygenation of indene (Allen 1997; Wackett et al. 1988). The oxygenation of indene in *Rhodococcus* strain I24 is followed by the oxidation of the indandiol products to keto-OH-indan. Generally, oxidations of this type are carried out by a diol dehydrogenase (Zylstra 1994). It has not been determined whether a single promiscuous dehydrogenase acts upon the different indandiols in *Rhodococcus* I24, or whether there are multiple stereospecific dehydrogenases in this system.

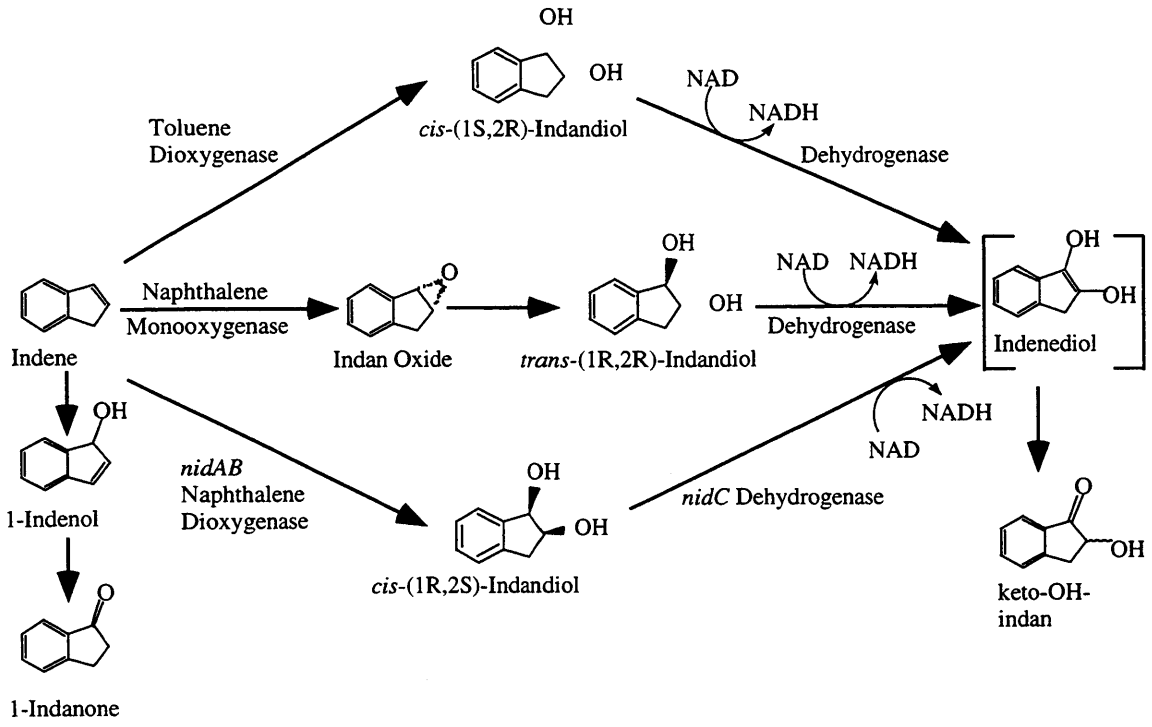


Figure 3-1: Proposed Indene Bioconversion Pathway in *Rhodococcus* strain I24
Genes encoding the naphthalene dioxygenase (*nidAB*) and a diol dehydrogenase (*nidC*) account for part of the indene bioconversion pathway. They were identified in a screen using a functional assay (see Chapter Two).

From a genomic library of *Rhodococcus* strain I24, we isolated genes encoding several enzymes involved in the bioconversion of indene including the large and small subunits of a naphthalene-type dioxygenase (*nidAB*), and a diol dehydrogenase (*nidC*). These genes were isolated on a single cosmid (pR4) that was identified via an indigo formation assay (Chapter Two)(Treadway et al. 1999). We demonstrated that when *nidAB* are expressed in a heterologous host indene is stereospecifically oxygenated to *cis*-(1R,2S)-indandiol. Furthermore, the *cis*-(1R,2S)-indandiol can be oxidized to keto-OH-indan when *nidC* is expressed. The NidC dehydrogenase is approximately 32% homologous to various diol dehydrogenases including BphB from *Rhodococcus* sp. strain RHA1, BphB from *Pseudomonas* sp. strain KKS102, and BphB from *Pseudomonas putida* OU83 (Treadway et al. 1999).

To better understand and control indene bioconversion in *Rhodococcus* I24 we wanted to further characterize the dehydrogenase encoded by *nidC* both genetically and biochemically. In this chapter we show through a variety of *in vitro* and *in vivo* assays that *nidC* encodes an enzyme that has diol dehydrogenase activity. Furthermore, we demonstrate that this activity is stereospecific for the substrate *cis*-(1R,2S)-indandiol. We also investigate the consequences of deleting this gene in the host strain *Rhodococcus* I24. We discuss the implications of these findings on the proposed model for indene bioconversion in this strain and strategies for manipulating the indene bioconversion.

Materials and Methods

Reagents

All strains used in this study are listed in Table 3-1. The plasmids used in this study are listed in Table 3-2. All chemicals were reagent grade and purchased from Sigma (St. Louis, MO) unless otherwise noted. The *cis*-(1R,2S)-dihydro-naphthalenediol was obtained from Fluka (Milwaukee, WI). The *cis*-(1R,2S)-indandiol and *trans*-(1R,2R)-indandiol were synthesized as described below. The *cis*-(1S,2R)-indandiol was a gift from M. Chartrain and colleagues at Merck Research Laboratories (Rahway, NJ). All media components were purchased from Difco (Detroit, MI) unless otherwise noted.

Table 3-1: Strains

Strains	*Description*	*Reference*
Rhodococcus I24	Converts indene to indandiol, orange colonies	Buckland et al., 1998
R. erythropolis SQ1	Easily transformable isolate of *R. erythropolis*	Qwan and Dabbs, 1993
Rhodococcus KY1-3	Spontaneous mutant of *Rhodococcus* I24 with modified indene bioconversion profile	K. Yanagimachi (in preparation)
Rhodococcus ST-1	*Rhodococcus* I24 Δ*nidC*	This work
Escherichia coli JM109	*endA1, recA1, gyrA96, thi, hsdR17*, (r_k-, m_k+), *relA1, supE44*, λ-, Δ(*lac-proAB*), [F', *traID36, proAB, lac*I^qZΔ*M15*]	Yanisch-Peron, C., Viera, J., and Messing, J., 1985
Escherichia coli S17-1	*recA*-, *res*-, *mod*-, proline auxotrophy, RP4 2-Tc:mu-km::Tn*7*, Kan^S, Tp^R, Sm^R	Simon et al., 1983
Escherichia coli MM294-4	Nalidixic acid resistant derivative of *E. coli* MM294 (Bachmann et al. 1987)	Eric Dabbs (personal communication)

DNA Manipulation and Transformation

All DNA restriction digests and manipulations were carried out using standard methods (Sambrook et al. 1989). Enzymes were purchased from New England Biolabs (Beverly, MA). Plasmid DNA was prepared using the Wizard Maxiprep Kit or Miniprep Kit from Promega (Madison, WI), or by boiling lysis (Sambrook et al. 1989). All gel purifications of DNA fragments were executed with the Gene Clean III Kit from Bio101 (Vista, CA) following the manufacturer's instructions. *Rhodococcus erythropolis* SQ1was transformed via electroporation as previously described (Treadway et al. 1999). *Escherichia coli* JM109 (Yanisch-Perron et al. 1985) was also transformed via electroporation (Sambrook et al. 1989).

Table 3-2: Plasmids

Plasmids	*Description*	*Reference*
pAL231	Amp^R, colE1 ori, *tsr* gene from pGM160 (Wohlleben et al., 1989) was ligated as an *Eco*RI-*Sma*I fragment into the *Eco*RI and *Eco*RV sites of a pCRScript (Strategene, La Jolla, CA) derivative	This study
pAL235	Kan^R, Amp^R, colE1 ori, RP4 mobility element in pCR2.1-Topo	This study
pAPE12	Kan^R, NG2 ori, PTRC and $lacI^q$ from pTRC99a	Guillouet et al., 1999
pCR2.1-Topo	Kan^R, Amp^R, colE1 ori	Invitrogen (Carlsbad, CA)
pCR-Script	Amp^R, colE1 ori	Strategene (La Jolla, CA)
pR4	Kan^R, NG2 ori, I24 genomic cosmid clone containing *nidABCD*	Treadway et al., 1999
pR4-10	Kan^R, NG2 ori, *Sau*3a pR4 subclone containing *nidAB*	Treadway et al., 1999
pST108	Kan^R, Amp^R, colE1 ori, the 1.1 kb *Xba*I-*Nsi*I fragment from pST112 containing the upstream and downstream regions of *nidC* inserted into pST114 cut with *Xba*I and *Nsi*I	This study
pST109	Kan^R, Amp^R, $Thio^R$, colE1 ori, the *tsr* gene from pAL231 on a 1.14 kb *Bsp120*I-*Pst*I fragment ligated into pST108 cut with *Not*I and *Nsi*I	This study
pST112	Kan^R, Amp^R, colE1, derivative of pCR-Topo2.1 with 1.1 kb PCR fragment of 5' upstream region of pR4 dehydrogenase *nidC* using primers DH-1 and DH-2	This study
pST114	Kan^R, Amp^R, colE1, derivative of pCR-Topo2.1 with 1.4 kb PCR product containing the 3' downstream region of the pR4 dehydrogenase *nidC* using primers DH-3 and DH-4	This study
pST116b	Kan^R, Amp^R, colE1, derivative of pCR-Topo2.1 using primers 5'-dehydro and 3'-dehydro to amplify pR4 dehydrogenase *nidC*	This study
pST117	Kan^R, Amp^R, colE1, the 1.0 kb *nidC* gene as an *Eco*RI fragment from pST116b inserted into pAPE12 digested with *Eco*RI	This study
pST119	Amp^R, $Thio^R$, colE1 ori, the RP4 mobility element on a *Pst*I-*Spe*I fragment from pAL235 inserted into pST109 cut with *Pst*I and *Xba*I	This study
pSUP301	Kan^R, Amp^R, pACYC177 with RP4 mobility element	Simon et. al., 1983

Plasmid Construction

For the polymerase chain reaction (PCR) (Saiki et al. 1986) *Taq* polymerase, *Taq* buffer and nucleotides were purchased from Boehringer Mannheim (Indianapolis, IN) and used according to the manufacturer's instructions. The cycle conditions used with each PCR reaction are noted with each construction. All PCR reaction parameters have a 5 minute 95°C pre-cycle and a 10 minute 72°C post-cycle unless otherwise stated. Sequences of the PCR primers used in this study can be found in Table 3-3. Primers were purchased from Gibco BRL (Grand Island, NY).

Table 3-3: PCR Primers

Primer Name	*Primer Sequence*
SLT-8	5'- GGAAGCGCCCCCTCCACGGTG-3'
5'-dehydro	5'- GGAATTCATGAGCATCATCCACAACGA-3'
3'-dehydro	5'- CCGAATTCGCCGTCGCTGTTGATGAT-3'
DH-1	5'-AGTACTTCTAGAGCATGCCGGAGAACATCGTCGA-3'
DH-2	5'- TACGTAATGCATGTAGGCTCCTCCGTCACATGATCA-3'
DH-3	5'- AGGAGCCTACATGCATTACGTAGCGGCCGCACCATCATCAACAGCGA-3'
DH-4	5'- CCTGTGTAACTTCTGACGCA-3'
DehKO-1	5'- GAACTCGGACGTTCGTCGGTAG-3'
DehKO-2	5'- GATTCGCTGGAGGACTACCTC-3'
inThio-1	5'- CTCGACCTCGATTCGTCAGTG-3'
inThio-2	5'- GACGTCCTCGATCAACGTTG-3'
MunI-mob	5'- CCAATTGATCCTTTTTGTCCGGTGT-3'
XbaI-mob	5'-GCTCTAGAGATCGTGGCATCACCGA-3'

The plasmid pAL231 was made by excising the *tsr* gene from pGM160 (Wohlleben et al. 1989) as an *Eco*RI-*Sma*I fragment and ligating it into the *Eco*RI and *Eco*RV sites of a pCRScript (Strategene, La Jolla, CA) derivative. The plasmid pAL235 was constructed as follows. The RP4 mob element was amplified from pSUP301 (Simon et al. 1983) using primers MunI-mob and XbaI-mob. 5 % Dimethyl sulfoxide (DMSO) was included in the PCR reaction mixture. The mixture was first heated to 94°C for five minutes and then cycled thirty times with each cycle comprising 1 min at 94°C, 4 min at 52°C, and 3 min at 72°C. A final incubation for 10 min at 72°C was included to ensure completion of the final extension products. The 1.0 kb product was cloned into the pCR2.1-Topo vector from the pCR2.1-Topo-TA Cloning Kit (Invitrogen, Carlsbad, CA) following the manufacturer's instructions. The insert in pAL235 is oriented such that the *Xba*I site from the XbaI-mob primer is adjacent to the *Xba*I site in the polylinker of the pCR2.1-Topovector.

pST116b was constructed via PCR using pR4 as the template DNA to amplify the 1.0 kb *nidC* dehydrogenase gene with primers 5'-dehydro and 3'-dehydro. The cycle profile was 30 cycles at 95°C for 1 minute, 50°C for 1 minute, and 72°C for 2 minutes. The gel-purified PCR product was cloned into the

pCR2.1-Topo vector from the pCR2.1-Topo-TA Cloning Kit (Invitrogen, Carlsbad, CA). pST117, the dehydrogenase expression construct, was made by ligating a 1.0 kb EcoRI fragment from pST116b that contains the *nidC* dehydrogenase into pAPE12 (Guillouet et al. 1999) that had been digested with *Eco*RI and treated with shrimp alkaline phosphatase (Boehringer Mannheim, Indianapolis, IN).

Construction of the dehydrogenase knock-out construct was carried out in a series of steps. A 1.1 kb region upstream of the *nidC* dehydrogenase was amplified by PCR using primers DH-1 and DH-2 using pR4 as a template. The PCR reaction profile included 30 cycles at 95°C for 30 sec, 55°C for 30 sec, and 72°C for 1 min. The PCR product was gel purified and ligated into pCR2.1-Topo as described above. The resulting plasmid was named pST112. A 1.4 kb fragment of DNA from the region downstream of the *nidC* was amplified using pR4 as the template and primers DH-3 and DH-4. DMSO (5%) was added to the PCR reaction. The cycle profile included 30 cycles at 95°C for 1 min, 47°C for 1 min, and 72°C for 2 minutes. The PCR product was gel-purified and ligated into pCR2.1-Topo as described above to form pST114. pST108 was constructed by inserting the 1.1 kb *Xba* I-*Nsi* I fragment from pST112 (containing the 5' upstream region) into pST114 digested with *Xba* I and *Nsi* I. The *tsr* (thiostrepton resistance) gene was inserted between the upstream and downstream regions, replacing the entire dehydrogenase open reading frame. The *tsr* gene was obtained as a 1.14 kb *Bsp*120I-*Pst* I fragment from pAL231, and inserted into pST108 cut with *Not* I and *Nsi* I. The 1.0 kb *Pst* I-*Spe* I fragment of pAL235 containing the RP4 mobility element was ligated to the 6.4 kb *Pst* I-*Xba* I fragment of pST109 to produce the final plasmid for making the knock-out, pST119.

Indandiol Biosynthesis and Purification

Strain *Rhodococcus* KY1-3 (K. Yanagimachi, in preparation) was used to synthesize *trans*-(1R,2R)-indandiol and strain *R. erythropolis* SQ1(pR4) was used to synthesize *cis*-(1R,2S)-indandiol. A VirTis Omni-Culture base (The VirTis Company, Gardiner, NY) with a New Brunswick 2.0 liter capacity vessel (New Brunswick Scientific Co., Inc., Edison, NJ) and custom-made head plate were used. The fermentation was conducted at 30 ± 1°C with an agitation speed of 1000 ± 5 rpm. The pH was monitored with a combination electrode (Cole-Parmer Instrument Co., Vernon Hills, IL) with a B&C Electronics controller (pH 7615, B&C Electronics, Milano, Italy), and maintained at 7.0 ± 0.5 by addition of sterile 2 N NaOH or 2 N HCl via a Masterflex variable speed modular peristaltic pump electrode (Cole-Parmer Instrument Co., Vernon Hills, IL). The filter-sterilized air feed was maintained at 1.0 liter/min with Cole-Parmer gas mass flow controllers electrode (Cole-Parmer Instrument Co., Vernon Hills, IL).

Medium K (Buckland et al. 1999) was used for the fermentation except that the glycerol was replaced with 20 g/L glucose (Fisher, Fair Lawn, NJ), the $FeSO_4 \cdot 7\ H_2O$ was omitted, and 150µg/ml kanamycin was added if appropriate. When the culture reached an OD_{600} of 15 (UVIKON 810 spectrophotometer, Kontron Instruments, Zurich, Switzerland) indene was continuously added through an air delivery system

that consisted of two air streams delivered through Teflon tubing: one air stream bubbled through 200 ml indene and then combined with another air stream to dilute the indene air concentration (Figure 3-2). The flow rates of each air stream were adjusted for a total flow rate of 1.0 L/min with approximately 150 PPM indene (measured using a PE Photovac 2020 photo-ionization air monitor, PE Photovac, Markham, Ontario, Canada, set with a response factor of 0.3).

Indene metabolite concentrations were monitored every 12 hours using the reverse-phase HPLC assay (see below). 1.0 ml samples from the fermentor were extracted with 3 ml MilliQ water and 6 ml HPLC grade isopropyl alcohol (Mallinckrodt Inc, Paris, KY), mixed, and centrifuged at 2455 x g. The supernatant was filtered using 0.2 µm PVDF 13 mm syringe filters (Alltech, Deerfield, IL) before HPLC analysis. The fermentation was stopped when the indandiol concentration reached approximately 1.0 g/L. The cells were centrifuged for 10 minutes at 15344 x g. The supernatant was filtered through a Gelman Sciences Suporcap-50 0.2 µm sterile filter (Gelman Sciences, Ann Arbor, MI) before purification. The indandiol in the supernatant was purified using the protocol of Chartrain and colleagues (Chartrain et al. 1998) except that ethanol was substituted for methanol as the eluant. Each indandiol obtained was at least 99% pure as determined by the chiral assay of Chartrain and colleagues (Chartrain et al. 1998).

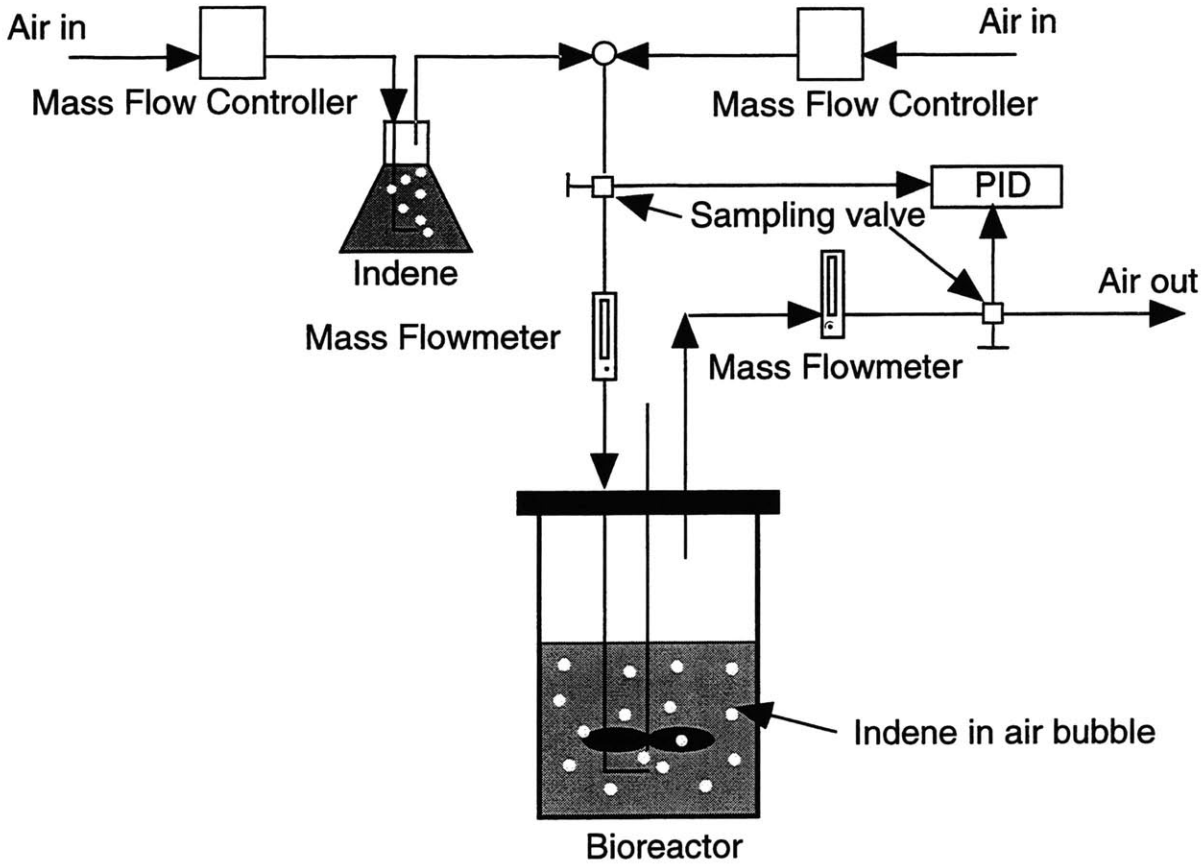
Preparation of cell lysate

A 200 ml LB culture with 150µg/ml kanamycin and 0.8 mM isopropyl-β-D-thiogalactopyranoside (IPTG) inoculated with *Rhodococcus* SQ1(pST117) cells was grown for two days to an OD_{600} of 3.0 (Spectronic 20 Genesys Spectrophotometer, Spectronic Instruments, Rochester, NY). Cells were harvested by centrifugation for 10 min at 4000 x g. Cells were then washed in 25 ml of PEG Buffer [0.5 M potassium phosphate buffer (pH 7.6),10% ethanol, 10% glycerol] and recentrifuged at 4000 x g for 10 minutes. Cells were resuspended in PEG buffer containing Complete Protease Inhibitor (Boehringer Mannheim, Indianapolis, IN) to an OD_{600} of 1000. The cell suspension was lysed using a mini-French Press cell (Spectronic Instruments NY, NY) set at 14,000 psi. The lysate was cleared via centrifugation at 47000 x g for 20 minutes. The protein concentration was determined using the Bradford assay using bovine serum albumin as the standard (Bio-Rad Laboratories Hercules, CA) following the manufacturer's instructions.

Dehydrogenase *in vitro* assay

Enzyme activity was determined by monitoring NADH production at 340 nm using a HP 8452A Diode Array Spectrophotometer (Hewlett Packard, Wilmington, DE). At its final concentration the reaction contained 0.1 mM potassium phosphate buffer (pH 7.6), 1.0 mM NAD^+, 10 mM diol substrate, and 340 µg total lysate protein in a volume of 1.0 ml. The diol substrates included: *cis*-(1R,2S)-indandiol, *cis*-(1S,2R)-indandiol, *cis*-(1R,2S)-dihydro-naphthalenediol, R-styrene glycol, and S-styrene glycol.

Figure 3-2: Indene Delivery System for Fermentations
This figure is courtesy K. Yanagimachi. Air is bubbled through a flask of indene and the indene saturated air is then fed to the fermentor. The indene was measured using a photo-ionization detector at entry and exit.



In vivo Dihydrodiol Breakdown Experiments

25 ml LB cultures of *R. erythropolis* SQ1 and derivative strains were grown for 12 hours at 30°C. The cultures contained 150 µg/ml kanamycin and/or 0.8mM isopropyl-β-D-thiogalactopyranoside (IPTG), if appropriate. For each culture, a 1.0 ml sample was taken. Subsequently, 2.5 mg of dihydrodiol was added directly to the medium and resuspended. Another 1.0 ml sample was taken immediately after resuspension. Samples were taken at regular intervals thereafter. To each sample 1.0 ml of HPLC grade isopropanol (Mallinckrodt, Paris, KY) was added. The cell suspension was vortexed and centrifuged for 5 min. Samples were filtered with 0.22 µm PVDF syringe filter (Alltech, Deerfield, IL) and analyzed via HPLC as described below. Analysis of *cis*-(1R,2S)-indandiol breakdown in *Rhodococcus* ST-1 and *Rhodococcus* I24 was performed as described above except that 10 µg/ml thiostrepton was added to the LB culture of *Rhodococcus* strain ST-1. Colony forming units (CFU/ml) were determined by plating serial dilutions of the cultures on LB plates supplemented with 10 µg/ml thiostrepton if appropriate. The plates were grown at 30°C and the colonies were counted after 2-3 days of growth.

R. erythropolis SQ1 Indene Bioconversions

R. erythropolis SQ1, *R. erythropolis* SQ1(pR4) and *R. erythropolis* SQ1(pR4-10) fermentations were carried out using the fermentation apparatus as described in the indandiol biosynthesis section. The fermentation medium was a defined medium called Medium Rare containing 40 g/L glucose (Fisher, Fair Lawn, NJ), 1.4 g/L $(NH_4)_2SO_4$, 1.0 g/L $MgSO_4 \cdot 7H_2O$, 0.015 g/L $CaCl_2 \cdot 2H_2O$ (Mallinckrodt Inc., Paris, KY), 1.0 g/L MOPS buffer, 1.0 ml/L A9 trace elements solution, 1.0 ml/L Stock A solution, 35.2 ml/L 1.0 M phosphate buffer, 150 µg/ml kanamycin (if appropriate), and 1.0 ml/L polypropylene glycol (MW 2000) (Acros Organics, NJ). The A9 trace elements solution contains 0.5 g/L $FeSO_4 \cdot 7H_2O$ (Mallinckrodt Inc., Paris, KY), 0.4 g/L $ZnSO_4 \cdot 7H_2O$ (Fluka, Buchs, Switzerland), 0.02 g/L $MnSO_4 \cdot H_2O$ (Mallinckrodt Inc., Paris, KY), 0.015 g/L H_3BO_3 (Fisher, Fair Lawn, NJ), 0.01 g/L $NiCl_2 \cdot 6H_2O$ (Mallinckrodt Inc., Paris, KY), 0.05 g/L $CoCl_2 \cdot H_2O$ (Mallinckrodt Inc., Paris, KY), 0.005 g/L $CuCl_2 \cdot 2H_2O$ (Mallinckrodt Inc., Paris, KY), and 0.25 g/L EDTA. The Stock A solution contains 2.0 g/L $NaMoO_4 \cdot 2H_2O$ and 5.0 g/L FeNa·EDTA (J.T. Baker, Phillipsburg, NJ). The 1.0 M phosphate buffer contains 113 g/L K_2HPO_4 and 47 g/L KH_2PO_4 (Mallinckrodt Inc., Paris, KY).

The fermentation and indene feed was initiated as described above. 1.0 ml samples were taken three times a day for three days during the indene feed. The samples were extracted and filtered as described above. Indene metabolite concentrations were measured using the reverse phase HPLC assay.

Reverse Phase HPLC Assay

An HPLC system (Waters 2690 Separation Module with Waters 996 Photodiode Array Detector, Waters Corporation, Milford, MA) equipped with a YMC CombiScreen C8 column (4.6 mm ID x 50 mm, S-

5 µm, 120 Å pore size) (YMC, Inc., Wilmington, NC) was used. Separation was achieved using a gradient-based elution with a mobile phase of acetonitrile (Mallinckrodt Inc, Paris, KY) and water delivered at a flow rate of 3.0 ml/min. The first two minutes of the HPLC were run isocratically at 90% water and 10% acetonitrile. The amount of acetonitrile was increased to 30% for the next minute and raised to 90% over the following minute. Compounds were detected at 220 nm at 23°C. The following elution times were obtained: *trans*-indandiol at 1.3 min, *cis*-indandiol at 1.5 min, 1-keto 2-hydroxy-indan at 1.8 min, 1-indenol at 3.2 min, 1-indanone at 3.5 min, and indene at 4.2 min. HPLC standards were gifts from Chartrain and colleagues at Merck Research Laboratories (Rahway, NJ).

Construction of *Rhodococcus* I24Δ*nidC*

Creation of a *nidC* null mutant was carried out by introducing the knock-out plasmid pST119 into *Rhodococcus* strain I24 via transconjugation using filters as previously described (Schafer et al. 1990). *E. coli* S17-1 (Simon et al. 1983) harboring pST119 was used as the donor strain, and *Rhodococcus* strain I24 was the recipient strain. Transconjugants were selected on LB plates supplemented with 10 µg/ml thiostrepton and 10 µg/ml nalidixic acid after 5-6 days at 30°C.

Genomic DNA was prepared from each Δ*nidC* candidate as described previously (Chapter Two)(Treadway et al. 1999). PCR reactions for analysis of candidate deletion strains included 5% DMSO. The following primers were used in this analysis (Table 3-3): SLT-8, DH-1, DH-4, DehKO-1, DehKO-2, inThio-1, and inThio-2. The cycle profile was 95°C for 5 minutes, followed by 30 cycles of 95°C for 1 minute, 49°C for 1 minute, and 72°C for 3 minutes. The 30 cycles were followed by a 10 minute extension cycle at 72°C. The PCR products were analyzed by electrophoresis on a 1% agarose gel. The PCR products were also analyzed by restriction digest using the enzymes *Sal*I and *Eco*RV.

Indene Bioconversions of *Rhodococcus* I24 and *Rhodococcus* ST-1

Uninduced indene bioconversions of *Rhodococcus* I24 and *Rhodococcus* ST-1 were carried out in a batch fermentations (apparatus described above). The conditions and sampling were the same as described above for the *Rhodococcus* SQ1 fermentations except that 10 µg/ml thiostrepton was added to the culture of *Rhodococcus* ST-1.

The naphthalene-induced bioconversions were conducted in screw capped flasks. A 500 ml culture in Medium Rare (see above) with 10 µg/ml thiostrepton, if appropriate, and 5 g/L of naphthalene was grown at 30°C until approximately an OD_{600} of 1.0. The culture was separated into three 100 ml aliquots into 500 ml screw capped flasks. Indene was added directly to the culture (no oil) in 20 µl aliquots, every 24 hours. 1.0 ml samples were taken every 12 hours for HPLC analysis. Samples were extracted with 1.0 ml of HPLC grade isopropanol (Mallinckrodt Inc., Paris, KY), filtered with 0.22 µm PVDF syringe filter (Alltech,

Deerfield, IL), and analyzed by HPLC as described above. 1.0 ml samples were taken every 24 hours to monitor the optical density.

Indene Bioconversion Induction Assay

500 ml Medium Rare (see above) cultures with 10 µg/ml thiostrepton, if appropriate, with and without 5 g/L of naphthalene were grown at 30°C for 43 hours. The cells were centrifuged for 10 minutes at 4000 x g. Cells were washed in induction buffer (10 g/L MOPS, 10 g/L glucose, 0.1 g/L $FeSO_4 \cdot 7H_2O$) and recentrifuged. The cells were resuspended in 40 ml of induction buffer in a 500 ml screw capped flask. 20 µl of indene was added directly to the flask. Samples (1.0 ml) were taken at regular intervals for two hours. 1.0 ml of HPLC grade isopropanol (Mallinckrodt Inc., Paris, KY) was immediately added to each sample. The samples were vortexed, filtered with 0.22 µm PVDF syringe filter (Alltech, Deerfield, IL), and analyzed by HPLC as described above.

Results

In vitro Dihydrodiol Enzyme Assay

We previously demonstrated that expression of *nidABC* promotes the conversion of indene to *cis*-(1R,2S)-indandiol, keto-OH-indan, 1-indenol, and 1-indanone (Treadway et al. 1999)(Chapter 2). Keto-OH-indan is believed to be the product of *cis*-(1R,2S)-indandiol dehydrogenation. To further characterize the NidC dehydrogenase thought to be responsible for this reaction we constructed an inducible plasmid, pST117, that would express only the dehydrogenase. To determine whether this gene encoded a functional dehydrogenase an *in vitro* enzyme assay was developed. For these assays we utilized a host strain, *R. erythropolis* SQ1, that is unable to metabolize indene or its derivatives. We transformed the plasmid pST117 into the naïve host *R. erythropolis* SQ1, grew the transformed *R. erythropolis* SQ1(pST117) cells in the presence of the inducer (IPTG), and prepared cleared crude lysates. We monitored the dehydrogenation reaction by recording the production of NADH using spectrophotometric analysis at 340 nm. We chose *cis*-(1R,2S)-dihydro-naphthalenediol as the substrate for the initial enzyme assays. This substrate was chosen because we believe that the dehydrogenase is associated with a naphthalene-type dioxygenase and that *cis*-(1R,2S)-dihydro-naphthalenediol may be the natural substrate of this enzyme. Naphthalene can be used as a carbon source for this strain, whereas indene can not (Chartrain et al. 1998).

Each enzyme assay reaction set contained the following: a blank with no lysate, a negative reaction with no substrate, and the positive reaction with substrate. We found that the NidC dehydrogenase displayed enzyme activity and can utilize *cis*-(1R,2S)-dihydro-naphthalenediol as a substrate (Figure 3-3 and Table 3-4). Using *cis*-(1R,2S)-dihydro-naphthalenediol as the substrate, standard curves for the enzyme activity with regards to substrate concentration and enzyme concentration (in this case total

protein from crude lysate) were determined (Figure 3-3). The effect of NAD^+ concentration was also determined (data not shown); the concentration of NAD^+ used in these assays is not limiting. All subsequent enzyme assays were based on this information. Since the dehydrogenase is not purified, enzyme activity is assessed on a plus or minus basis in all assays henceforth. Multiple assays were performed to show that the assay is reproducible.

In order to test the proposed model of indene bioconversion (Figure 3-1), we investigated the substrate specificity of the dehydrogenase in regards to the different indandiols produced by *Rhodococcus* I24 during an indene bioconversion. Specifically, we tested the substrates *cis*-(1S,2R)-indandiol and *cis*-(1R,2S)-indandiol. The *trans*-(1R,2R)-indandiol was unavailable in quantities large enough for this experiment. The results are shown in Table 3-4. We found that the NidC dehydrogenase could utilize *cis*-(1R,2S)-indandiol, but not *cis*-(1S,2R)-indandiol. To investigate the stereospecificity further we tested whether R- or S-styrene glycol could act as a substrate for the dehydrogenase in our enzyme assay. Neither of these compounds acted as a substrate for the dehydrogenase (Table 3-4). The observation that the NidC dehydrogenase can utilize *cis*-(1R,2S)-indandiol and *cis*-(1R,2S)-dihydro-naphthalenediol suggests that NidC prefers its substrates in the *cis*-(1R,2S) configuration.

Table 3-4: *In vitro* Dehydrogenase Enzyme Assay

Substrate	Specific Activity	
	SQ1(pST117) + Dehydrogenase	SQ1(pAPE12) -- Dehydrogenase
cis-(1R,2S)-indandiol	0.0037 ± 0.0016	--[a]
cis-(1S,2R)-indandiol	--	--
cis-(1R,2S)-dihydro-naphthalenediol	0.0416 ± 0.0125	--
R-styrene glycol	--	Nd[b]
S-styrene glycol	--	Nd[b]

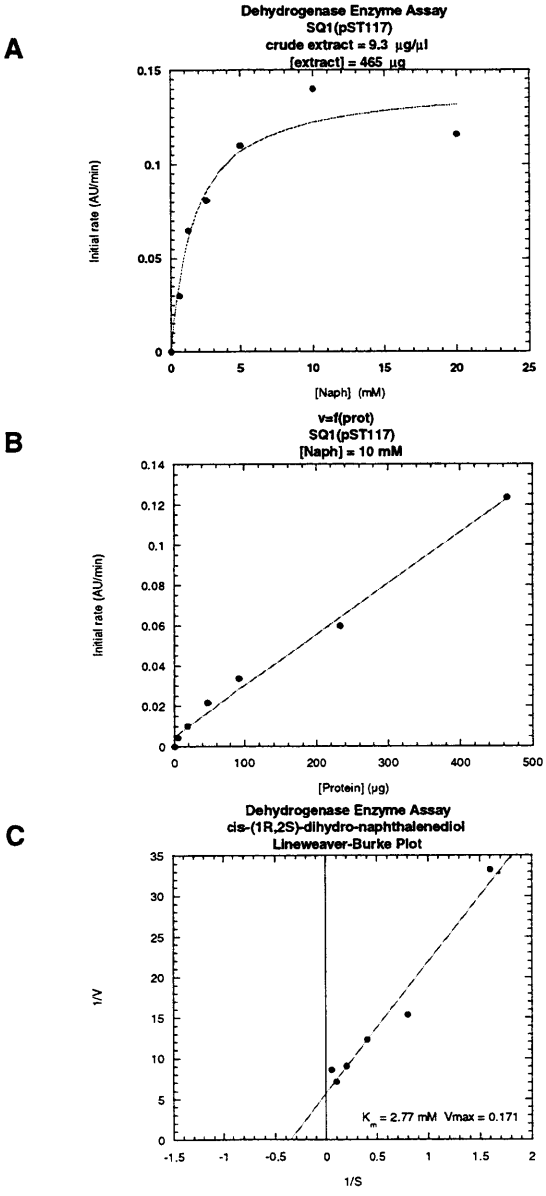
[a] -- = < 0.0001
[b] Nd, not determined

In vivo Dihydrodiol Breakdown

Since we had determined that the *nidC* gene encoded a dehydrogenase activity *in vitro*, we wanted to investigate the *in vivo* characteristics of the dehydrogenase. This type of analysis would provide insights into how the indene bioconversion pathway in *Rhodococcus* I24 might be manipulated. Therefore, we developed an *in vivo* diol breakdown assay. This assay requires the addition of a given diol substrate directly to the medium of a growing culture. Samples of the culture are then withdrawn over time and analyzed by HPLC to determine whether the diol is still present in the medium or if it has been converted to the dehydrogenated product.

Figure 3-3: Dehydrogenase Enzyme Assay Standard Curves with *cis*-(1R,2S)-dihydro-naphthalenediol

To assess the rate of the dehydrogenation reaction and the reproducibility of the assay, standard curves using *cis*-(1R,2S)-dihydro-naphthalenediol as the substrate were developed. A) Initial rate as a function of substrate concentration, B) Initial rate as a function of total protein concentration, C) Lineweaver-Burke Plot.



A series of *R. erythropolis* SQ1 derivative strains were analyzed for their ability to breakdown various diols *in vivo*. These strains include *R. erythropolis* SQ1 transformed with: the original cosmid clone pR4 containing the dioxygenase subunits and the dehydrogenase; a cosmid subclone only containing the dioxygenase subunits (pR4-10); the dehydrogenase expression plasmid (pST117); and the control plasmid containing no inserted genes (pAPE12). The results are shown in Table 3-5. The *in vivo* results mirror the *in vitro* analysis and confirm that the NidC dehydrogenase is stereospecific for the *cis*-(1R,2S)-diols. That is, *in vivo* the dehydrogenase can oxidize both *cis*-(1R,2S)-indandiol and *cis*-(1R,2S)-dihydro-naphthalenediol, but not *cis*-(1S,2R)-indandiol. Furthermore, the dehydrogenase does not oxidize *trans*-(1R,2R)-indandiol. We observed that the *cis*-(1R,2S)-dihydro-naphthalenediol was broken down in the presence of the dehydrogenase. However, the breakdown product was undetectable by HPLC and may have been further metabolized by the host strain *R. erythropolis* SQ1.

Table 3-5: *In vivo* Diol Breakdown Assay

Substrate	*SQ1(pAPE12) vector only*	*SQ1(pR4) cosmid clone*	*SQ1(pR4-10) dioxygenase only*	*SQ1(pST117) dehydrogenase only*
cis-(1R,2S)-indandiol	--	+	--	+
cis-(1S,2R)-indandiol	--	--	--	--
trans-(1R,2R)-indandiol	--	--	--	--
cis-(1R,2S)-dihydro-naphthalenediol	--	+[a]	--	+[a]

[a] Substrate was consumed, product was undetectable by HPLC

Indene Bioconversion by *Rhodococcus* SQ1 derivative strains

Since the indene bioconversion pathway in *Rhodococcus* I24 is a complex network (Figure 3-1), we wanted to begin to simplify the system in order to gain a better understanding of the pathway. One approach is to construct a system in which the components of interest are added to the exclusion of others. The naïve host *R. erythropolis* SQ1 harboring the plasmids pR4 and pR4-10 represents one such system. Based on the DNA sequence information (Treadway et al. 1999) we know that pR4 contains the wild type dioxygenase genes and the dehydrogenase gene. We also know that pR4-10 contains the dioxygenase genes but only the first 66 amino acids (out of 339) from the dehydrogenase. The truncated dehydrogenase lacks the putative NAD^+ binding site and the Y-x-x-K putative active site motif (Jornvall et al. 1995; Treadway et al. 1999). Consequently, the subclone pR4-10 is the genetic equivalent of a dehydrogenase null mutation.

We carried out indene bioconversions in batch fermentations with *R. erythropolis* SQ1(pR4) and *R. erythropolis* SQ1(pR4-10) to see if there was a difference in the product formation profiles. These fermentations were executed in the absence of silicon oil as a carrier for the indene. Instead, indene was delivered in the vapor phase in closed vessels (a fermentor in this instance). The oil was removed because *Rhodococcus* strains tend to grow exclusively in the oil phase and cling to the walls of the vessel; *R. erythropolis* SQ1 is particularly adept at this. Other strains of *Rhodococcus* have demonstrated this behavior as well (Neu 1996). The fermentation results are shown in Figure 3-4. No products were formed in the *R. erythropolis* SQ1 control fermentation (data not shown). The major phenotypic difference between *R. erythropolis* SQ1(pR4) (Figure 3-4A) and *R. erythropolis* SQ1(pR4-10) (Figure 3-4B) is the lack of keto-OH-indan production by *R. erythropolis* SQ1(pR4-10). *R. erythropolis* SQ1(pR4-10) behaves as expected for a *nidC* null mutant strain.

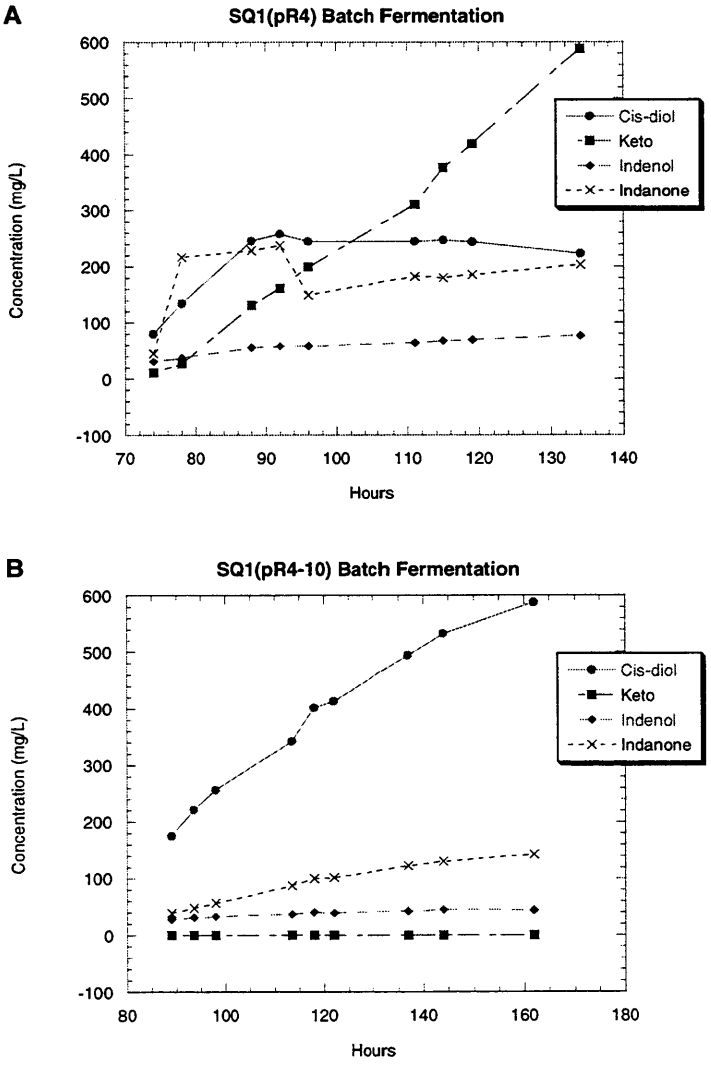
Deletion of the *nidC* Dehydrogenase Gene in *Rhodococcus* I24

Since the fermentation analysis suggested a clear phenotype for the *nidC* dehydrogenase deletion mutant, we wanted to carry out a similar analysis of the dehydrogenase gene in the parent strain *Rhodococcus* I24. We introduced the knock-out plasmid pST119 into *Rhodococcus* I24 by transconjugation with *Escherichia coli*. Since this plasmid is unable to replicate in *Rhodococcus*, we selected thiostrepton resistant colonies to identify strains in which the plasmid had undergone homologous recombination to completely replace the *nidC* dehydrogenase gene with the *tsr* gene. As a control, *E. coli* S17-1(pST119) was mated to *E. coli* MM294-4, a nalidixic acid resistant derivative of *E. coli* MM294 (Bachmann 1987). Transconjugants were selected on 100 μg/ml ampicillin and 10 μg/ml nalidixic acid, and grown at 37°C overnight. Plasmid DNAs isolated from *E. coli* MM294-4 transconjugants were analyzed by restriction enzyme digestion and agarose gel electrophoresis to confirm presence of the plasmid pST119 (data not shown). Using this methodology we produced *Rhodococcus* I24Δ*nidC*, called *Rhodococcus* ST-1 henceforth.

We confirmed the presence of the gene replacement and its location in the genome by PCR analysis of genomic DNA. A schematic of this analysis using five sets of primers is depicted in Figure 3-5. Figure 3-6 shows the PCR results generated using genomic DNA isolated from the *nidC* deletion strain *Rhodococcus* ST-1, *Rhodococcus* I24 , and plasmid DNA controls. The DNA

Figure 3-4: A model of the *nidC* knock-out: Indene Bioconversion of *Rhodococcus* SQ1(pR4) and *Rhodococcus* SQ1(pR4-10)

Indene bioconversions were carried out in a batch fermentor to determine the product profile that would result in a simulation of a *nidC* dehydrogenase deletion. A) *Rhodococcus* SQ1(pR4) represent the wild-type dehydrogenase and B) *Rhodococcus* SQ1(pR4-10) represents the dehydrogenase knock-out. The optical density of the fermentation cultures was similar. No *trans*-indandiol is produced by *Rhodococcus* SQ1(pR4) or *Rhodococcus* SQ1(pR4-10). The *Rhodococcus* SQ1 fermentation did not produce indene bioconversion product.



fragments amplified from the genomic DNA of the knock-out strain *Rhodococcus* ST-1 were the sizes predicted for a deletion mutant. Primer set one produced a few extra products with the cosmid template pR4 that are presumably due to mis-priming of the primers on the template DNA. To further demonstrate that the PCR products resembled the expected products they were digested with selected enzymes (data not shown). All of the expected products digested predictably. None of the additional products from primer set one with the pR4 template digested with the appropriate enzymes.

After confirming the presence of the gene deletion, wild-type *Rhodococcus* I24 and *Rhodococcus* ST-1 strains were analyzed in a series of experiments to determine the effect of removing the *nidC* dehydrogenase. The first set of experiments included uninduced and naphthalene-induced indene bioconversions. The purpose of these experiments was to examine the overall phenotypic effect of deleting the *nidC* dehydrogenase on the indene bioconversion profile. If this dehydrogenase is a major contributor to the dehydrogenase activity in this system, removing this activity should have an effect on the accumulation of *cis*-indandiol and the production of keto-OH-indan. The results from these experiments are shown in Figures 3-7 and 3-8. There was no significant difference in the indene bioconversion product profiles of *Rhodococcus* I24 (wild type) and *Rhodococcus* ST-1 either uninduced or naphthalene-induced.

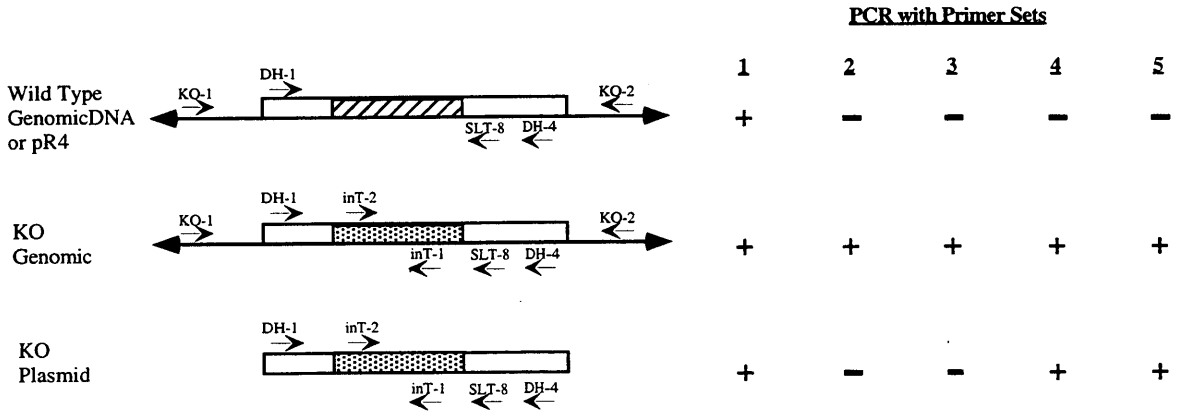
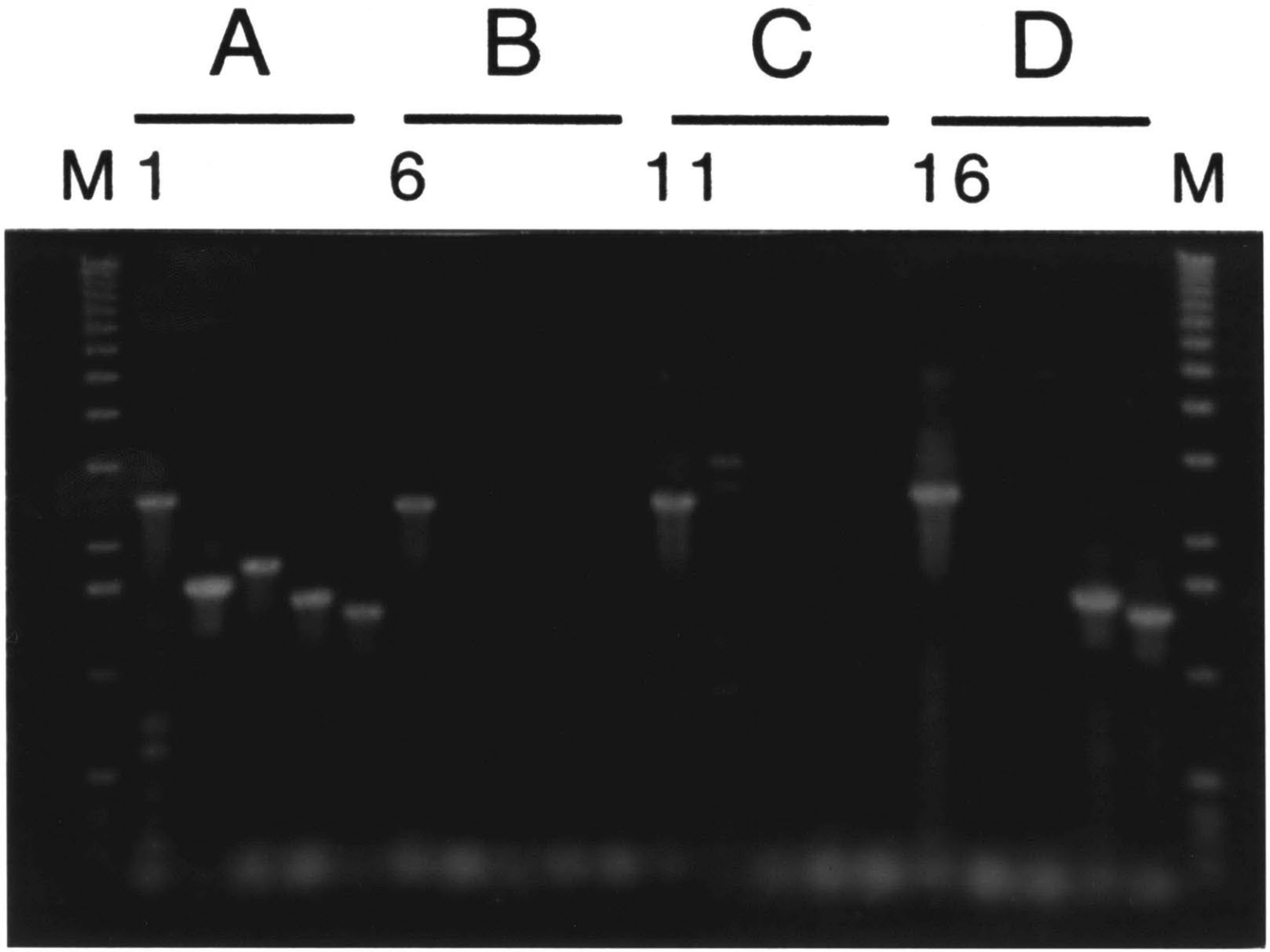


Figure 3-5: Schematic of PCR Confirmation of *Rhodococcus* I24Δ*nidC*
Depiction of the PCR results expected using primers designed to confirm the presence and location of the thiostrepton gene in *Rhodococcus* I24Δ*nidC*. Sequences of the primers can be found in Table 3-3. The primer sets are as follows: primer set 1 DH-1 and SLT-8, primer set 2 DehKO-1 and inThio-1, primer set 3 DehKO-2 and inThio-2, primer set 4 inThio-2 and DH-4, primer set 5 inThio-1 and DH-1. For shorthand purposes on the diagram DehKO-1 is KO-1, DehKO-2 is KO-2, inThio-1 is inT-1, and inThio-2 is inT-2. The hatched box is the *nidC* gene and the stippled box is the *tsr* gene.

Figure 3-6: PCR Analysis of the *nidC* knock-out *Rhodococcus* ST-1 (next page)
PCR analysis was carried out genomic DNA and plasmid DNA to confirm the presence of the *nidC* dehydrogenase knock-out in *Rhodococcus* I24. Lanes 1-5 are ST-1 genomic DNA with primer sets 1,2,3,4,and 5 respectively. Lanes 6-10 are wild-type I24 genomic DNA with primer sets 1,2,3,4,and 5 respectively. Lanes 11-15 are pR4 plasmid DNA with primer sets 1,2,3,4,and 5 respectively. Lanes 16-20 are pST119 knock-out plasmid DNA with primer sets 1,2,3,4,and 5 respectively. The primer sets are as indicated in figure 3-5.


A
B
C
D
M 1
6
11
16
M

The lack of a global phenotype alteration in the long term indene bioconversion experiments with *Rhodococcus* ST-1 indicated that closer inspection was necessary. It is possible that the effect of the dehydrogenase knock-out is negated in the long term by a second activity. Therefore, a difference may be seen in the early stages of an indene bioconversion but not the late stages. Consequently, we examined the induction profiles of the wildtype strain and the *nidC* null mutant in the first two hours of an indene bioconversion. The results of this analysis are shown in Figure 3-9. There was no difference in the naphthalene induction profiles of *Rhodococcus* I24 and *Rhodococcus* ST-1. The uninduced cells did not display indene bioconversion in the two hour time frame examined (data not shown).

Since we had determined that the NidC diol dehydrogenase was stereospecific for *cis*-(1R,2S)-indandiol, an *in vivo* diol breakdown assay using *cis*-(1R,2S)-indandiol as the substrate was conducted as the third set of experiments. We found a significant rate difference in the oxidation of *cis*-(1R,2S)-indandiol between the wildtype and the null mutant strains (Figure 3-10A and 3-10B). The Δ*nidC* strain, *Rhodococcus* ST-1, oxidizes the *cis*-(1R,2S)-indandiol one third as rapidly as the wild-type strain. Furthermore when *Rhodococcus* I24 and *Rhodococcus* ST-1 cells are pre-induced with naphthalene, the rate of *cis*-(1R,2S)-indandiol breakdown is the same (Figure 3-11C and 3-11B). However, when the rates of *cis*-(1R,2S)-indandiol breakdown in *Rhodococcus* ST-1 uninduced and naphthalene-induced cells are compared, the there is a significant rate difference (Figure 3-11A and 3-11B). The uninduced *Rhodococcus* ST-1 cells have a much slower rate of *cis*-(1R,2S)-indandiol oxidation than the naphthalene-induced cells. No rate differences were found when *cis*-(1S,2R)-indandiol or *cis*-(1R,2S)-dihydro-naphthalenediol were used as substrates (data not shown).

Figure 3-7: Indene bioconversion of *Rhodococcus* I24 and *Rhodococcus* ST-1
Indene bioconversions were conducted in batch fermentations in collaboration with K. Yanagimachi. A) *Rhodococcus* ST-1, B) *Rhodococcus* I24, C) Optical Density (OD 600) of cultures.

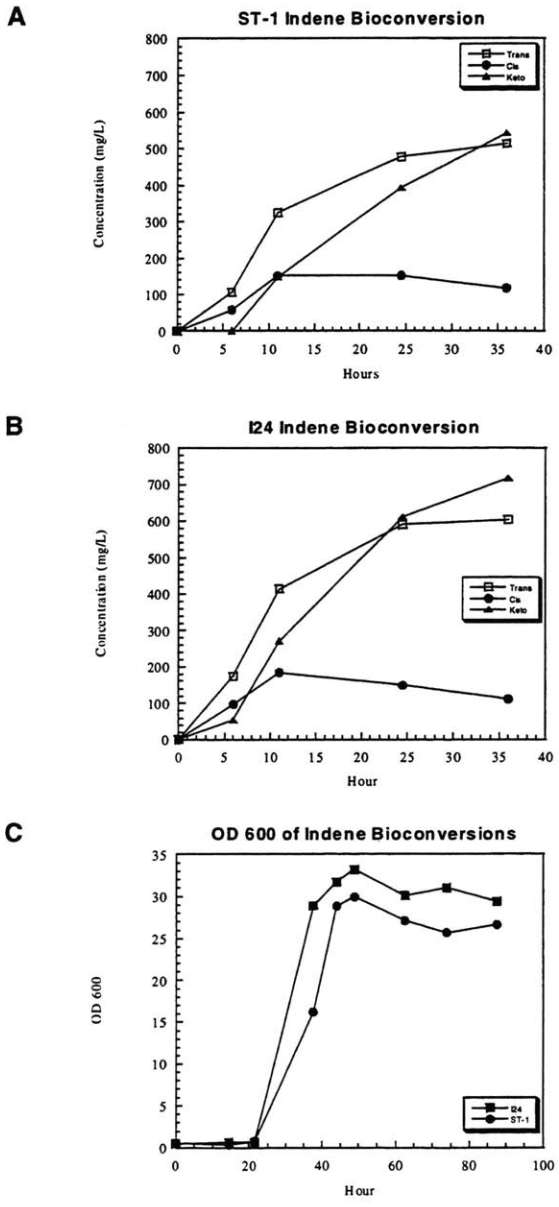


Figure 3-8: Naphthalene-induced indene bioconversion by *Rhodococcus* I24 and *Rhodococcus* ST-1

These experiments were conducted in screw capped flasks. There is some variability in these experiments when conducted in a shake flask. The key feature of these results is the overall trend and final relative positions of the indene metabolites produced. A) *Rhodococcus* ST-1 naphthalene-induced, B) *Rhodococcus* I24 naphthalene-induced, C) Optical Density (OD 600) of cultures.

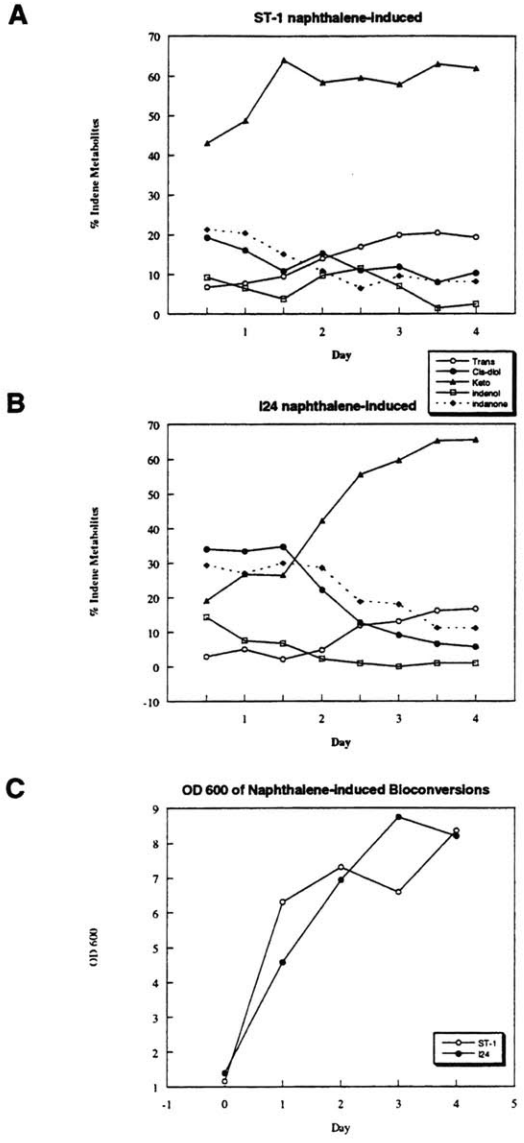


Figure 3-9: Two hour induction analysis of *Rhodococcus* I24 and *Rhodococcus* ST-1
Cultures of *Rhodococcus* I24 and *Rhodococcus* ST-1 were grown in the presence and absence of naphthalene. Indene was added at time zero and samples were taken at regular intervals to assess indene bioconversion activity. A) *Rhodococcus* I24 naphthalene-induced, B) *Rhodococcus* ST-1 naphthalene-induced. The uninduced cultures (not shown) showed no indene bioconversion activity in the first two hours.

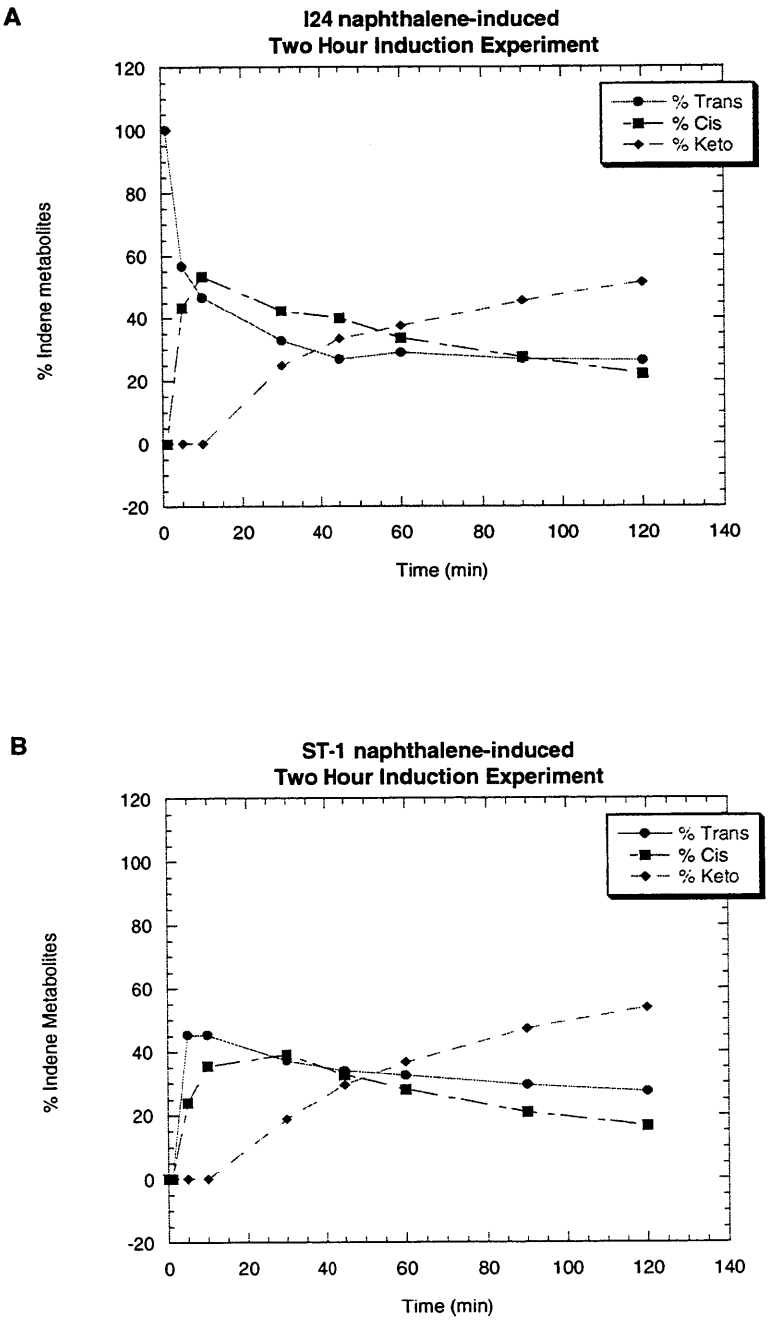
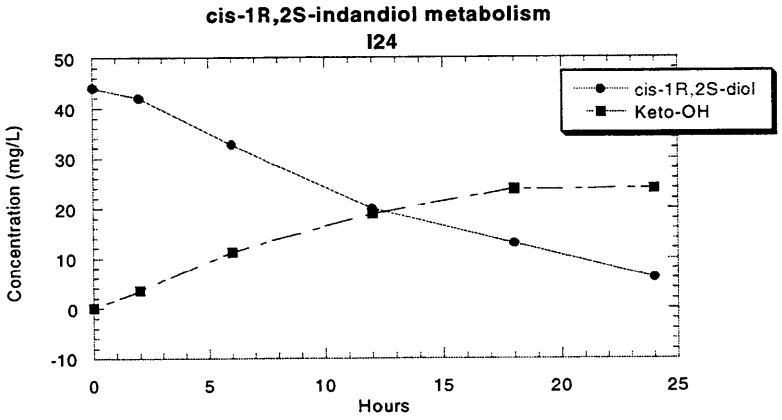


Figure 3-10: *cis*-(1R,2S)-indandiol Breakdown by *Rhodococcus* I24 and *Rhodococcus* ST-1 *cis*-(1R,2S)-indandiol was added directly to cultures of *Rhodococcus* ST-1 (*Rhodococcus* I24Δ*nidC*) and *Rhodococcus* I24. The breakdown of the *cis*-(1R,2S)-indandiol and the production of keto-OH-indan was monitored via HPLC analysis. A) *Rhodococcus* I24 *cis*-(1R,2S)-indandiol breakdown, B) *Rhodococcus* ST-1 *cis*-(1R,2S)-indandiol breakdown. CFU/ml was measured and had no effect on the interpretation of the results when incorporated.

A



B

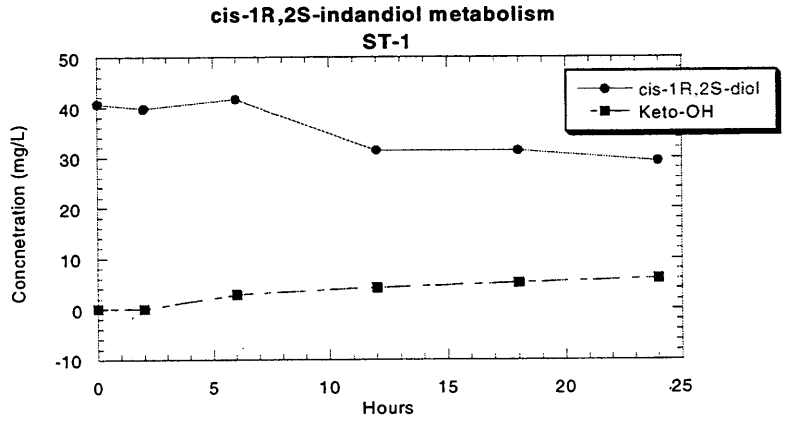
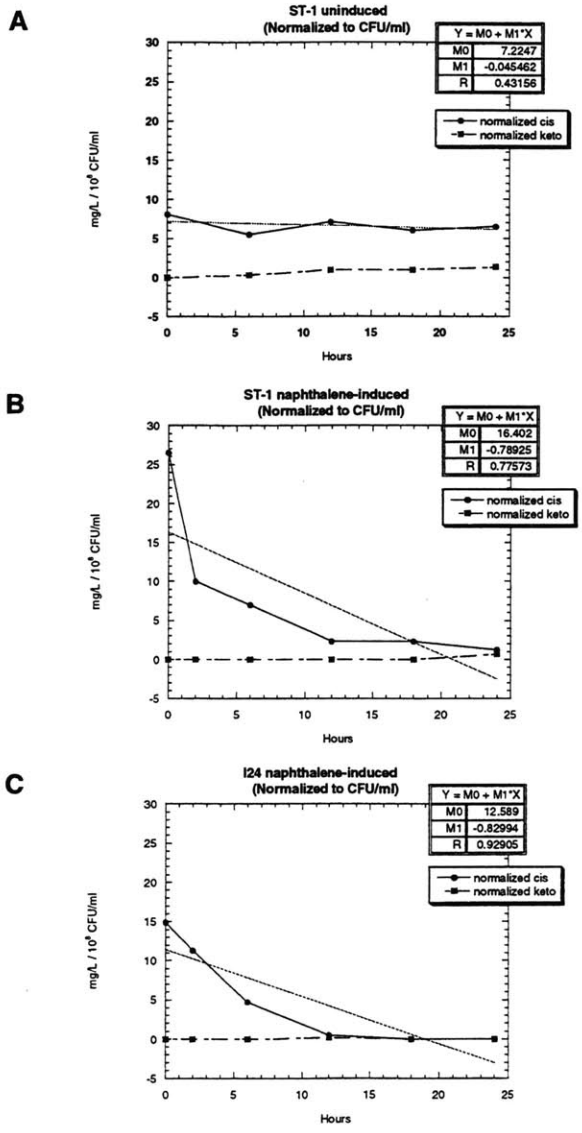


Figure 3-11: Naphthalene-induced *cis*-(1R,2S)-indandiol Breakdown by *Rhodococcus* ST-1 Cultures were pre-grown in naphthalene for 24 hours prior to adding *cis*-(1R,2S)-indandiol. The breakdown of the *cis*-(1R,2S)-indandiol and the production of keto-OH-indan was monitored via HPLC analysis. All graphs are normalized to CFU/ml since culture conditions were different in the presence versus the absence of naphthalene. A) uninduced *Rhodococcus* ST-1 *cis*-(1R,2S)-indandiol breakdown, B) naphthalene-induced *Rhodococcus* ST-1 *cis*-(1R,2S)-indandiol breakdown, C) naphthalene-induced *Rhodococcus* I24 *cis*-(1R,2S)-indandiol breakdown.



Discussion

In our earlier work we identified the *nidC* gene and showed it was homologous to a variety of diol dehydrogenases (Chapter Two)(Treadway et al. 1999). In this work we have demonstrated using both biochemical and genetic analysis that the *nidC* gene encodes a functional diol dehydrogenase. *In vivo* characterization of the *nidC* gene expressed in a heterologous host demonstrated that the *nidC* gene product could oxidize dihydrodiol compounds. Furthermore, the *in vitro* enzyme assay results mirror the *in vivo* results. This supports the notion that NidC is a diol dehydrogenase.

In our analysis of the NidC dehydrogenase we found that it has two levels of selectivity. First as demonstrated in the *in vivo* analysis, the dehydrogenase discriminates between *cis* and *trans* forms of indandiol. Other diol dehydrogenases have been shown to be substrate selective in this way (Fong et al. 1996; Patel and Gibson 1974; Rogers and Gibson 1977). Second, the NidC dehydrogenase is stereospecific for a certain *cis*-indandiol enantiomer. In two separate assays the dehydrogenase could utilize *cis*-(1R,2S)-indandiol as a substrate, but not *cis*-(1S,2R)-indandiol. This degree of stereospecificity has been demonstrated in *Pseudomonas* and other organisms (Allen et al. 1995; Buckland et al. 1999; Eaton et al. 1996; Jeffrey et al. 1975; Sato et al. 1994). Other diol dehydrogenases have been shown to have a preference for one enantiomer over the other, but are still able to oxidize both forms (Burczynski et al. 1998; Geerlof et al. 1994). Our results also suggest that this specificity may be determined by the presence of the -OH group at the 2 position since neither enantiomer of styrene glycol could be utilized as a substrate by the NidC dehydrogenase. It is interesting to note that the NidC dehydrogenase is specific for the *cis*-indandiol that is produced by the NidAB dioxygenase that is genetically associated with the dehydrogenase. This suggests the possibility that each oxygenase system in *Rhodococcus* I24 (Figure 3-1) has a coordinating stereospecific dehydrogenase activity. Based on the stereospecificity of the NidC dehydrogenase, we conclude that there is at least one more diol dehydrogenase activity in this strain.

To better understand the indene bioconversion pathway in *Rhodococcus* I24 we simplified the system by removing the dehydrogenase activity. We did this in two ways: by analyzing the *nidABC* system in isolation and by studying *Rhodococcus* I24 lacking *nidC*. The first method utilized the heterologous host *R. erythropolis* SQ1, a related species unable to metabolize indene. We constructed *R. erythropolis* SQ1 derivative strains carrying the dioxygenase and dehydrogenase (*nidABC*) or carrying the dioxygenase only (*nidAB*). In this system removing the *nidC* dehydrogenase prevented the oxidation of *cis*-(1R,2S)-indandiol in an indene bioconversion. This result was further confirmed by the *in vivo* diol breakdown assay. When we applied the same analysis to *Rhodococcus* I24 and *Rhodococcus* ST-1 (*Rhodococcus* I24Δ*nidC*) we found a more complicated scenario. We observed a reduction in the amount of *cis*-(1R,2S)-indandiol dehydrogenation rather than abolishment of the activity in uninduced cells. Consequently, it appears that the *nidC* dehydrogenase is expressed at low levels without inducing the system with indene

or naphthalene. Moreover, our results show that there is at least one other diol dehydrogenase in *Rhodococcus* I24 capable of utilizing *cis*-(1R,2S)-indandiol as a substrate. When *Rhodococcus* ST-1 is pre-induced with naphthalene the rate of *cis*-(1R,2S)-indandiol oxidation is the same as in the naphthalene-induced wild-type strain *Rhodococcus* I24. This suggests that this additional dehydrogenase activity may be naphthalene-inducible. Based on our proposed pathway (Figure 3-1), this activity could be the dehydrogenase associated with the naphthalene-inducible monooxygenase. Therefore, it is possible that the other dehydrogenase(s) in the *Rhodococcus* I24 indene bioconversion pathway is(are) not stereospecific. However, it is also possible that all of the keto-OH-indan is derived from the *cis*-(1R,2S)-indandiol. We are in the process of investigating the effects of *nidC* overexpression in *Rhodococcus* I24 especially in terms of the ratio of *cis*-(1R,2S)-indandiol to *cis*-(1S,2R)-indandiol.

Rhodococcus I24 employs three oxygenase systems and at least two diol dehydrogenases in the conversion of indene to keto-OH-indan. We have shown that one of these dehydrogenases is stereospecific for one of the indandiols produced in this network, specifically *cis*-(1R,2S)-indandiol. As we begin to unravel the indene bioconversion pathway in *Rhodococcus* I24 we find increasing layers of complexity. We are in the process of isolating the genes encoding the enzymes in the other branches of this pathway so we may further understand the indene bioconversion network. As we continue to study bacteria capable of stereospecific bioconversions we hope to develop a controllable process for the production of chiral compounds. At the same time we will gain a fundamental understanding of how *Rhodococcus*, and other organisms like it, metabolize aromatic hydrocarbons.

Acknowledgements

This work was supported by a grant from Merck Research Laboratories. S. L. T. was supported in part by NIH Biotechnology Training Grant # 2T32 GM08334-10.

We would like to thank Merck Research Laboratories for their support. We also thank members of the Sinskey Lab for their discussion, comments and support. We especially thank S. Guillouet for technical assistance, K. Yanagimachi for assistance with fermentations and HPLC analysis, and L. B. Willis for editorial comments.

References

Allen CCR, D. R. Boyd, M. J. Larkin, K. A. Reid, N. D. Sharma, K. Wilson (1997) Metabolism of Napthalene, 1-Napthol, Indene, and Indole by *Rhodococcus* sp. Strain NCIMB 12038. Appl Environ Microbiol 63: 151-155

Allen CCR, Boyd DR, Dalton H, Sharma NDB, I., Kerley NA, Sheldrake GN, Taylor SC (1995) Enantioselective bacterial biotransformation routes to *cis*-diol metabolites of monosubstituted

benzenes, naphthalene and benzocycloalkenes of either absolute configuration. J. Chem. Soc. Chem. Commun. 1995: 117-118

Bachmann BJ (1987) *Escherichia coli* and *Salmonella typhimurium*, Cellular and Molecular Biology. ASM

Boyd DR, McMordie RAS, Sharma ND, Dalton H, Williams P, Jenkins RO (1989) Stereospecific benzylic hydroxylation of bicyclic alkenes by *Pseudomonas putida*: Isolation of (+)-R-1,2-dihydronaphthalene, an arene hydrate of naphthalene from metabolism of 1,2-dihydronaphthalene. J Chem Soc, Chem Commun : 339-340

Buckland B, Drew S, Connors N, Chartrain M, Lee C, Salmon P, Gbewonyo K, Gailliot P, Singhvi R, Olewinski R, Sun W, Reddy J, Zhang J, Zhou W, Jackey B, Goklen K, Junker B, Greasham R (1999) Microbial conversion of Indene to Indandiol, a key intermediate in the synthesis of CRIXIVAN. Metabolic Engineering 1: 63-74

Burczynski ME, Harvey RG, Penning TM (1998) Expression and characterization of four recombinant human dihydrodiol dehydrogenase isoforms: oxidation of trans-7, 8-dihydroxy-7,8-dihydrobenzo[a]pyrene to the activated o-quinone metabolite benzo[a]pyrene-7,8-dione. Biochemistry 37: 6781-90

Chartrain M, Jackey B, Taylor C, Sandford V, Gbewonyo K, Lister L, DiMichele L, Hirsch C, Heimbuch B, Maxwell C, Pascoe D, Buckland B, Greasham R (1998) Bioconversion of indene to *cis*-(1S,2R)-indandiol and *trans*-(1R,2R)-indandiol by *Rhodococcus* species. J Ferment Bioeng 86: 550-558

Collins AN, G.N. Sheldrake, and J. Crosby (1997) Chirality in Industry II: Developments in Manufacturing and Applications of Optically Active Compounds. John Wiley & Sons, Inc., New York

Eaton SL, Resnick SM, Gibson DT (1996) Initial reactions in the oxidation of 1,2-dihydronaphthalene by *Sphingomonas yanoikuyae* strains. Appl Environ Microbiol 62: 4388-94

Fong KP, Goh CB, Tan HM (1996) Characterization and expression of the plasmid-borne *bedD* gene from *Pseudomonas putida* ML2, which codes for a NAD+-dependent cis-benzene dihydrodiol dehydrogenase. J Bacteriol 178: 5592-601

Geerlof A, Jongejan JA, van Dooren TJ, Racemakers-Franken PC, van den Tweel WJ, Duine JA (1994) Factors relevant to the production of (R)-(+)-glycidol (2,3-epoxy-1- propanol) from racemic glycidol by enantioselective oxidation with *Acetobacter pasteurianus* ATCC 12874. Enzyme Microb Technol 16: 1059-63

Guillouet S, Rodal AA, An G-H, Lessard PA, Sinskey AJ (1999) Overexpression of the *E. coli* isoleucine feedback resistant catabolic threonine dehydratase in *Corynebacterium glutamicum* and its effect on isoleucine production. Appl Environ Microbiol submitted

Hamberg M, Zhang LY, Brodowsky ID, Oliw EH (1994) Sequential oxygenation of linoleic acid in the fungus *Gaeumannomyces graminis*: stereochemistry of dioxygenase and hydroperoxide isomerase reactions. Arch Biochem Biophys 309: 77-80

Jeffrey AM, Yeh HJ, Jerina DM, Patel TR, Davey JF, Gibson DT (1975) Initial reactions in the oxidation of naphthalene by *Pseudomonas putida*. Biochemistry 14: 575-84

Jornvall H, Persson B, Krook M, Atrian S, Gonzalez-Duarte R, Jeffery J, Ghosh D (1995) Short-chain dehydrogenases/reductases (SDR). Biochemistry 34: 6003-13

Lee K, Sol M. Resnick, David T. Gibson (1997) Stereospecific Oxidation of (R)- and (S)-1-Indanol by Naphthalene Dioxygenase from *Pseudomonas* sp. Strain NCIB 9816-4. Appl Environ Microbiol 63: 2067-2070

Neu TR (1996) Significance of bacterial surface-active compounds in interaction of bacteria with interfaces. Microbiol Rev 60: 151-66

Patel RN (1997) Stereoselcetive Biotransformations in Syntheis of Some Pharmaceutical Intermediates. In: Neidleman SL, Laskin AI (eds) Advacnec in Applied Micribiology, vol 43. Academic Press, San Diego, pp 91-140

Patel TR, Gibson DT (1974) Purification and properties of (+)-cis-naphthalene dihydrodiol dehydrogenase of *Pseudomonas putida*. J Bacteriol 119: 879-88

Resnick SM, Daniel S. Torok, Kyoung Lee, John M. Brand, David T. Gibson (1994) Regiospecific and Stereoselective Hydroxylation of 1-Indanone and 2-Indanone by Napthalene Dioxygenase and Toluene Dioxygenase. Appl Environ Microbiol 60: 3323-3328

Rogers JE, Gibson DT (1977) Purification and properties of cis-toluene dihydrodiol dehydrogenase from *Pseudomonas putida*. J Bacteriol 130: 1117-24

Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA (1986) Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. Nature 324: 163-6

Sambrook J, Fritsch E, Maniatis T (1989) Molecular Cloning: A Laboratory Manual, 2nd edition edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sato K, Nakanishi M, Deyashiki Y, Hara A, Matsuura K, Ohya I (1994) Purification and characterization of dimeric dihydrodiol dehydrogenase from dog liver. J Biochem (Tokyo) 116: 711-7

Schafer A, Kalinowski J, Simon R, Seep-Feldhaus AH, Puhler A (1990) High-frequency conjugal plasmid transfer from gram-negative *Escherichia coli* to various gram-positive coryneform bacteria. J Bacteriol 172: 1663-6

Simon R, Priefer U, Puhler A (1983) A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram Negative bacteria. Biotechnology Nov 1983: 784-791

Treadway SL, Yanagimachi K, Lankenau E, Lessard PA, Stephanopoulos G, Sinskey AJ (1999) Isolation and characterization of indene bioconversion genes from *Rhodococcus* strain I24. Appl Microbiol Biotechnol In press

Wackett LP, Kwart LD, Gibson DT (1988) Benzylic monooxygenation catalyzed by toluene dioxygenase from *Pseudomonas putida*. Biochemistry 27: 1360-7

Wohlleben W, Arnold W, Bissonnette L, Pelletier A, Tanguay A, Roy PH, Gamboa GC, Barry GF, Aubert E, Davies J, et al. (1989) On the evolution of Tn21-like multiresistance transposons: sequence analysis of the gene (*aacC1*) for gentamicin acetyltransferase-3- I(AAC(3)-I), another member of the Tn21-based expression cassette. Mol Gen Genet 217: 202-8

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-19

Zylstra GJ (1994) Molecular Analysis of Aromatic Hydrocarbon Degradation. In: Garte SJ (ed) Molecular Environmental Biology. Lewis Publishers, Boca Raton, pp 83-115

Chapter Four: Conclusion

Summary

Rhodococcus strain I24 was recently isolated from soil contaminated with aromatic hydrocarbons. As a newly isolated strain, *Rhodococcus* strain I24 was genetically uncharacterized. The diverse nature of the *Rhodococcus* genus requires the of testing of different available genetic tools to determine if they are useful for analysis of the strain in question. Consequently, multiple genetic tools were analyzed and developed to facilitate analysis of *Rhodococcus* strain I24. In work not explicitly described here, numerous plasmids were obtained from other researchers to find suitable vectors to use in a variety of genetic analyses. One of these plasmids was used in the construction of the multi-host cosmid vector pRhodoCOS described in Chapter Two. In addition, numerous molecular biology protocols for plasmid isolation, genomic DNA isolation, transformation, and conjugation were examined for their effectiveness in *Rhodococcus* strain I24. Many of these molecular biology/microbiology techniques required considerable optimization whereas others had to be developed *de novo*. Without these tools this project would never have been possible.

Rhodococcus strain I24 is capable of utilizing aromatic hydrocarbons as carbon sources, specifically naphthalene and toluene. The bioconversion of indene to three enantiomers of indandiol, keto-OH-indan, 1-indenol and 1-indanone is an interesting feature of this strain. Two of the indandiols produced can be used as a precursor for the HIV-1 protease inhibitor indinavir (Crixivan™) (see Chapter One). Cloning of the different genes encoding the enzymes responsible for this bioconversion has led to an enhanced understanding of the indene bioconversion network in *Rhodococcus* strain I24. A functional screen identified genes encoding a ring-hydroxylating dioxygenase system (*nidAB*) that catalyzes the conversion of indene to *cis*-(1R,2S)-indandiol. Based on the induction studies of Chartrain and colleagues (Chartrain et al. 1998), the *cis*-(1R,2S)-indandiol is produced from indene by a naphthalene-inducible dioxygenase. Consequently, the NidAB dioxygenase could be naphthalene inducible; however, this has never been demonstrated experimentally. Analysis of naphthalene-induced indene bioconversions in *R. erythropolis* SQ1(pR4), a strain expressing the *nid* genes on a plasmid, could possibly answer this question. If the appropriate regulatory factors are present on the pR4 clone there could be a dramatic increase in the production of *cis*-(1R,2S)-indandiol in naphthalene-induced cells as compared to uninduced cells. Another way to address the issue of naphthalene inducibility would be to integrate a reporter gene, such as *lacZ*, into the genome replacing the dioxygenase and monitor β-galactosidase activity when the cells are induced with naphthalene.

Researchers have demonstrated that electron transport proteins are required for dioxygenase activity (Ensley and Gibson 1983; Haigler and Gibson 1990a; Haigler and Gibson 1990b; Subramanian et al.

1985). Sequence analysis of the *nidAB* dioxygenase system indicated that there are no identifiable genes encoding the electron transport proteins genetically associated with the dioxygenase subunits. As discussed in Chapter Two, there may be novel genes for the electron transport or they may be borrowed from the host strain used in the experiments. Alternatively, NidAB may represent a new class of dioxygenase system not previously described. An indene bioconversion using a strain containing a plasmid expressing just the dioxygenase subunits from a known promoter could experimentally address this. If the construct promotes the bioconversion of indene, then the dioxygenase subunits are either borrowing the electron transport proteins or they are not required. The dioxygenase subunits may be unable to promote indene bioconversion. Therefore the smallest functional unit of the subclone pR4-10, which is known to express the dioxygenase subunits, could be determined via nested deletion analysis. There are five kb of DNA upstream of the dioxygenase subunits in this subclone that could contain novel genes for the reductase and ferredoxin electron transport proteins. Unfortunately, the possibility of borrowing electron transport proteins can never really be fully eliminated regardless of the host strain used.

In addition to genes encoding dioxygenase subunits, a diol dehydrogenase (*nidC*) was also identified. The gene encoding this enzyme is located next to the dioxygenase small subunit gene. This enzyme is responsible for the conversion of the *cis*-(1R,2S)-indandiol to keto-OH-indan, as demonstrated in indene bioconversion experiments. Further analysis of the diol dehydrogenase, described in Chapter Three, determined that the enzyme is stereospecific for *cis*-(1R,2S)-indandiol produced by the *nidAB* dioxygenase system. It is not capable of using any other indandiol produced during a *Rhodococcus* strain I24 indene bioconversion as a substrate. The *nidC* null mutant strain is still capable of oxidizing *cis*-(1R,2S)-indandiol, although at a reduced rate. The implications of these findings are discussed below.

The *nidABC* genes appear to be a functional unit, although it remains to be determined if they are part of a true operon. Often operons containing genes encoding aromatic hydrocarbon metabolism pathways are located on plasmids. Both the naphthalene and toluene metabolism genes in *Pseudomonas putida* are located on plasmids (Williams and Worsey 1976; Yen and Gunsalus 1982). Some of these plasmids have been found to be linear (Dabrock et al. 1994; Masai et al. 1997), and/or transmissible (Dabrock et al. 1994; Nakazawa and Yokota 1977; Shields et al. 1995). Genome analysis of *Rhodococcus* strain I24 determined that the genome is approximately 3.0 Mb and contains a large plasmid (pI24) that is about 340 kb. It would be interesting to determine if the *nidABC* genes are located on pI24, and if the plasmid pI24 is linear or transmissible. *Rhodococcus* strain I24 DNA separated by pulse field gel electrophoresis could be analyzed via Southern blotting could determine if the *nid* genes are chromosomally or plasmid encoded. To determine the structure of pI24, PFG samples of could be treated with exonuclease III and topoisomerase I. Exonuclease II treated DNA should exhibit a mobility shift or smearing due to digestion of the linear DNA. DNA samples treated with topoisomerase I would exhibit a mobility shift if the plasmid is circular due to nicking of the plasmid DNA. Nicking the DNA relaxes the supercoils of the plasmid.

Alternatively, the gene location and transmissibility of the plasmid could be determined simultaneously. If the *nid* genes are located on the plasmid it is possible that transfer of the plasmid to a naïve host could confer the ability to grow on aromatic carbon sources such as toluene or naphthalene. PFG analysis of the new strain would demonstrate transmission of the plasmid and genomic DNA analysis using homology based methods or PCR would demonstrate the presence of the *nid* genes. This analysis could be extremely useful as it could lead to the identification of other genes involved in the indene bioconversion pathway.

The natural substrate of the Nid enzymes is not known. We hypothesize that naphthalene may be a natural substrate since it can be used as a carbon source by the strain. However, we have never tested if naphthalene can act as a substrate for the NidAB dioxygenase. We do know that a naphthalene derivative, *cis*-(1R,2S)-dihydro-naphthalenediol, can act as a substrate for the NidC dehydrogenase. It would be interesting to see what compounds would be produced in the dioxygenase and dehydrogenation reactions when naphthalene is used as a substrate. A possible naphthalene degradation pathway is shown in Figure 4-1. One experiment to address this would be to use *R. erythropolis* SQ1(pR4), *R. erythropolis* SQ1(pR4-10), and *R. erythropolis* SQ1 in a naphthalene bioconversion experiment. The products produced could be isolated via HPLC separation and analyzed by NMR to determine their structures. The results expected from this experiment are described in Table 4-1 below. In Chapter Two we hypothesized that the NidAB dioxygenase is a naphthalene-type dioxygenase, the conversion of naphthalene to 1,2-dihydro-naphthalenediol could support that notion. Naphthalene utilization could be examined in *Rhodococcus* strain I24 and *Rhodococcus* ST-1 as well. Due to the presence of multiple oxygenase pathways, two of which appear to be naphthalene-inducible, it is likely that the bioconversion of naphthalene will produce multiple enantiomers of 1,2-dihydro-naphthalenediol as well as other compounds. Analysis of the naphthalene breakdown products in these two systems could provide insights into the indene bioconversion pathway and the stereospecificity of the enzymes involved.

Ring-hydroxylating dioxygenases can oxygenate in a stereospecific manner that may be related to the substrate. The NidAB dioxygenase stereospecifically oxygenates indene to *cis*-(1R,2S)-indandiol and 1-indenol. Other dioxygenase have been shown to stereospecifically oxygenate indene (Allen et al. 1995; Allen et al. 1997; Gibson et al. 1995; Wackett et al. 1988). Interestingly, the naphthalene dioxygenase (NDO) from *Pseudomonas putida* can oxygenate indene to *cis*-(1R,2S)-indandiol and

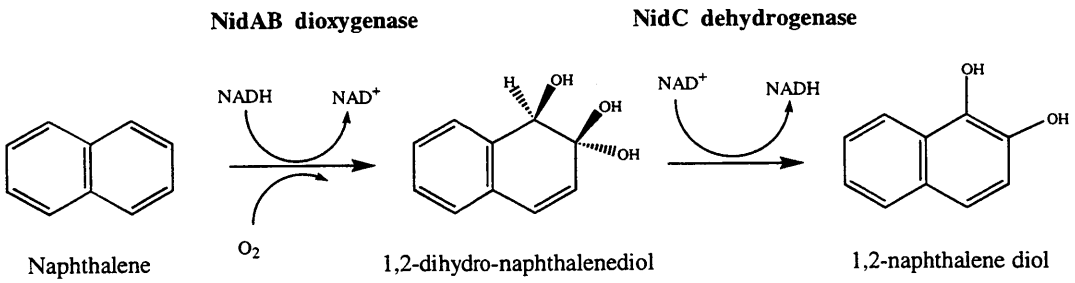


Figure 4-1: Proposed pathway of naphthalene oxygenation by NidABC System
A possible pathway for naphthalene oxygenation by the Nid enzymes is shown above. The pathway resembles a portion of the upper pathway of naphthalene degradation from *Pseudomonas putida*.

Table 4-1: Naphthalene as a substrate of NidABC enzymes

Strain	*Enzyme Content*	*Compounds Expected from Naphthalene*
R. erythropolis SQ1	none	none
R. erythropolis SQ1(pR4)	dioxygenase (large and small subunits)	cis-1,2-dihydro-naphthalenediol
	dehydrogenase	1,2-naphthalenediol
R. erythropolis SQ1(pR4-10)	dioxygenase (large and small subunits)	cis-1,2-dihydro-naphthalenediol

1S-indenol (Resnick et al. 1996). Based on sequence alignments, the NidAB α subunit is 30 % similar to the α subunit of the NDO from *P. putida*. Further comparison of the two dioxygenases based on data from the crystal structure of the NDO (Kauppi et al. 1998), suggests that NDO and NidA (the large subunit of the NidAB dioxygenase) may have a similar active site/mononuclear iron binding site. The active site residues of NDO have been suggested to be His 208, His 213, and Asp 362. The residue Asp 205 is also required for activity and likely is responsible for electron transfer (Parales et al. 1999). The NidA dioxygenase subunit aligns with the NDO and the strongest regions of identity are in the Reiske iron-sulfur center and the mononuclear iron binding/active site. In NidA the iron coordinating residues are likely His 216, His 221, and Asp 372. The Asp residue is spaced slightly differently, there are two extra residues between the His 221 and Asp 372. There is also a corresponding Asp residue, Asp 213, to the electron transfer Asp residue of NDO. It is unclear how the active sites of these enzymes play a role in substrate specificity.

New Models for Indene Bioconversion in I24

A new model for the indene bioconversion pathway is shown in Figure 4-2 based on all of the data described in this work. *Rhodococcus* strain I24 contains one dioxygenase system (NidAB) that is specific for the production of *cis*-(1R,2S)-indandiol. This dioxygenase also is responsible, at least partially, for producing 1-indenol and 1-indandone in an indene bioconversion. Coordinately expressed with the dioxygenase system is a diol dehydrogenase called NidC. The dehydrogenase oxidizes the *cis*-(1R,2S)-indandiol produced in an indene bioconversion, and only the *cis*-(1R,2S)-indandiol, to keto-OH-indan in this pathway. No other indandiol in this system can act as a substrate for the NidC dehydrogenase. Analysis of the *nidC* diol dehydrogenase null mutant strain, *Rhodococcus* ST-1, uncovered a second diol dehydrogenase activity in this strain that uses *cis*-(1R,2S)-indandiol as a substrate. This activity appears to be induced by naphthalene; consequently this dehydrogenase could be a dehydrogenase that is coordinately expressed with the naphthalene inducible monooxygenase activity. These data suggest that the indene bioconversion network is complex, and has multiple oxygenases and multiple dehydrogenases.

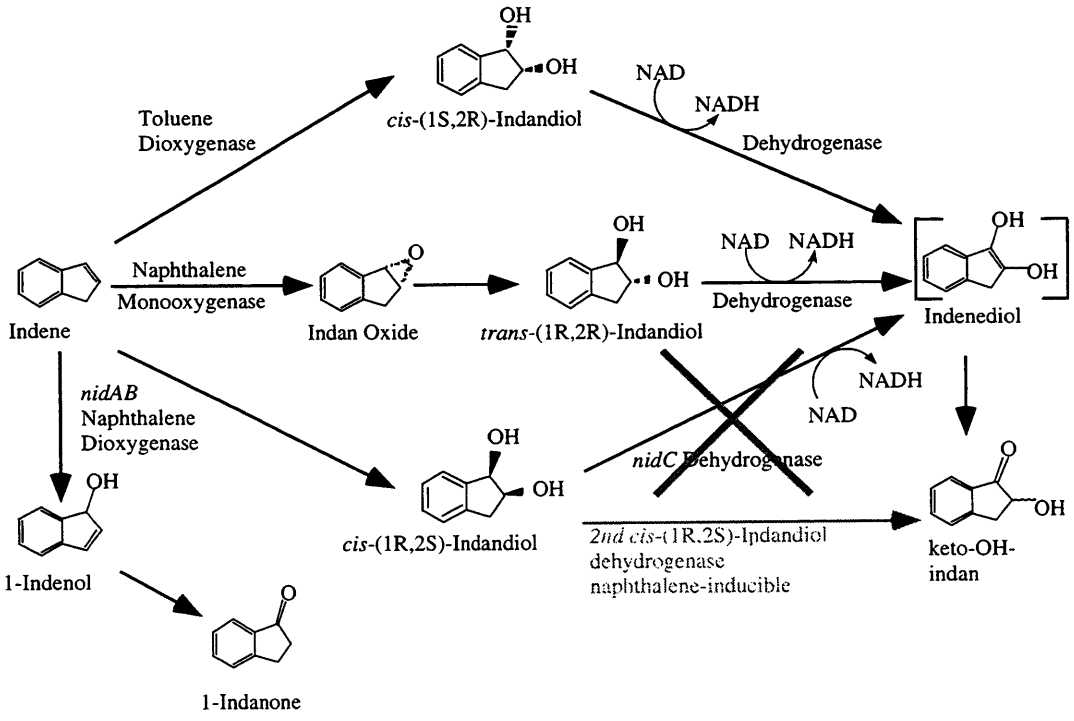


Figure 4-2: Modified Pathway for Indene Bioconversion Pathway in *Rhodococcus* I24
Above is the new proposed pathway for indene bioconversion including data from the *nidC* dehydrogenase knock-out strain, *Rhodococcus* ST-1. The analysis of the strain suggests that there is a second dehydrogenase capable of utilizing cis-(1R,2S)-indandiol as a substrate in *Rhodococcus* strain I24.

Work by others in the laboratory not discussed in this thesis also indicates that a greater degree of complexity in the indene bioconversion pathway than previously believed. Some of this data is discussed

below and a second model for the indene bioconversion pathway is proposed. The comprehensive model combining the data described in this thesis with data from other reasearch is shown in Figure 4-3. Two experiments were conducted that provided insights into the indene bioconversion pathway. Both experiments were designed to investigate the monooxygenation reaction and its components. *Rhodococcus* KY-1, a *Rhodococcus* strain I24 derivative strain that has an altered indene bioconversion profile, was used in these experiments. The first experiment was a biochemical analysis indan oxide and its breakdown during an indene bioconversion. The second experiment was a radiolabeled substrate utilization experiment using the *trans*-(1R,2R)-indandiol. The biochemical analysis of indan oxide demonstrated that there is a spontaneous breakdown of this intermediate in a 1:1 fashion to produce *trans*-(1R,2R)-indandiol and *cis*-(1S,2R)-indandiol. This was shown by monitoring the breakdown of chemically synthesized indan oxide in cleared cell lysates of *Rhodococcus* KY-1. Even when the cells were pre-induced with naphthalene or toluene the indan oxide still resolved into *trans*-(1R,2R)-indandiol and *cis*-(1S,2R)-indandiol in a 1:1 ratio. Consequently the indan oxide could predominantly be a single enantiomer, specifically R-indan oxide (A. Dexter, unpublished results). In order to account for the naphthalene induction studies of Chartrain and colleagues, there could be a naphthalene inducible epoxide hydrolase (or equivalent activity) in *Rhodococcus* strain I24 that is not in the derivative strain. Such an enzymatic activity could account for the increased production of the *trans*-(1R,2R)-indandiol over *cis*-(1S,2R)-indandiol in a naphthalene-induced indene bioconversion. However, it seems unlikely that such a hydrolase exists in this system even though the identity of the modification of *Rhodococcus* strain I24 that created *Rhodococcus* KY-1 is unknown. A more plausible explanation, that is depicted in the model Figure 4-3, is that there is a naphthalene-inducible dehydrogenase that can oxidize the *cis*-(1S,2R)-indandiol but not the *trans*-(1R,2R)-indandiol. The substrate utilization experiments using radiolabeled *trans*-(1R,2R)-indandiol provided some information towards this hypothesis. From these experiments it is believed that the *trans*-(1R,2R)-indandiol is not oxidized very rapidly, or at all, by a dehydrogenase in *Rhodococcus* strain I24 (K. Yanagimachi, unpublished results). This observation can also explain the overwhelming accumulation of *trans*-(1R,2R)-indandiol in long term (100 + hour) indene bioconversion fermentations of *Rhodococcus* strain I24 (Chartrain et al. 1998).

Both of these models are intriguing and suggest further experiments. Each of the models predict the presence of a second dehydrogenase. Experimental data presented in Chapter Three supports the first model suggesting the presence of a naphthalene-inducible dehydrogenase that can oxidize *cis*-(1R,2S)-indandiol . The second model also proposes the presence of a naphthalene-inducible dioxygenase, but this dehydrogenase can oxidize *cis*-(1S,2R)-indandiol. It is possible that these dehydrogenases may be one in the same, that is, a non-stereospecific dehydrogenase. Most dehydrogenases identified to date do not exhibit absolute stereospecificity. Instead the enzymes display preferences for one enantiomer as determined by rate differences in substrate utilization. To determine if there is naphthalene-inducible dehydrogenase that can oxidize *cis*-(1S,2R)-indandiol, radiolabeled substrate experiments could be

carried out using *Rhodococcus* strain I24 or *Rhodococcus* ST-1, either strain should produce the same results. The radiolabeled substrate improves quantitative analysis and reduces the quantity of substrate needed for the experiment. Using cells induced with naphthalene or uninduced, the breakdown

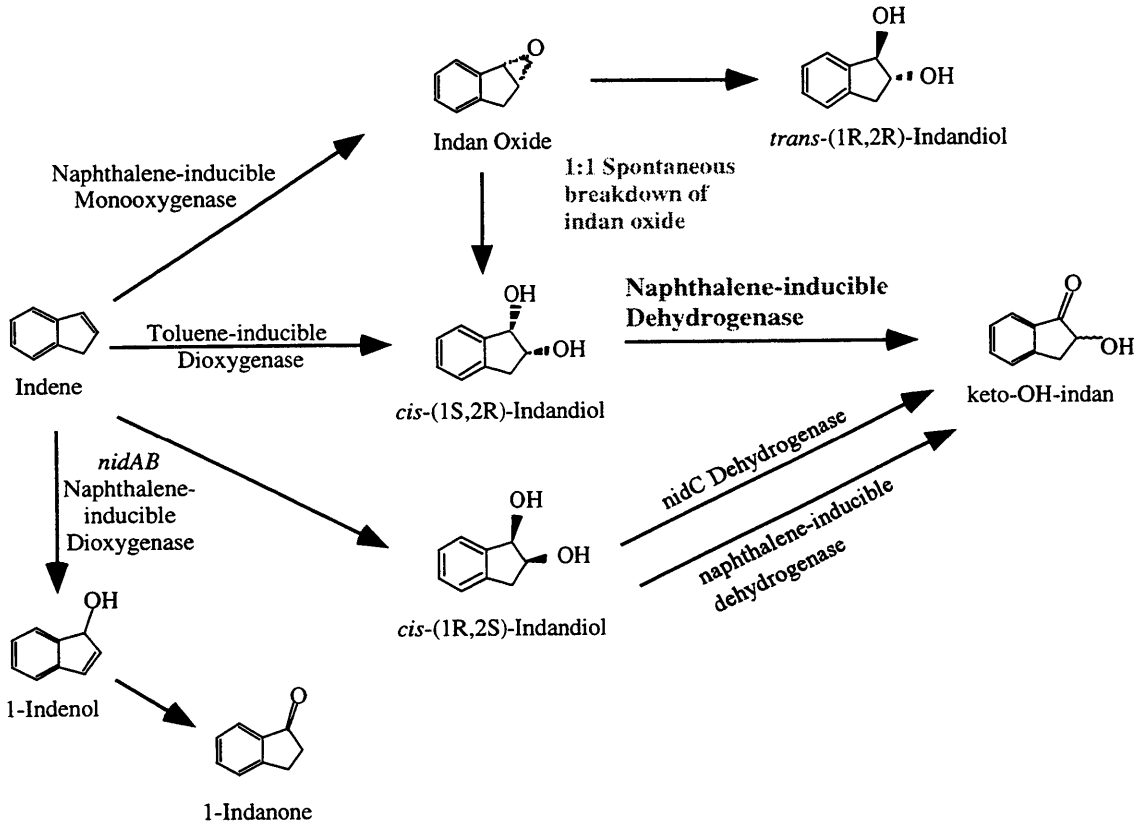


Figure 4-3: Comprehensive Model for Indene Bioconversion in *Rhodococcus* strain I24
This model is based on unpublished data from members of the Sinskey Laboratory and the Stephanopoulos Laboratory. In an indene bioconversion the indan-oxide appears to break down in a 1:1 ratio into *trans*-(1R,2R)-indandiol and *cis*-(1S,2R)-indandiol, suggesting a spontaneous reaction. Radiolabeled *trans*-(1R,2R)-indandiol does not rapidly breakdown during an indene bioconversion.

of radiolabeled *cis*-(1S,2R)-indandiol could be monitored by HPLC. The rate of *cis*-(1S,2R)-indandiol breakdown should be faster in the naphthalene-induced cells if there is a naphthalene-inducible dehydrogenase capable of using *cis*-(1S,2R)-indandiol as a substrate. Using information from the naphthalene bioconversion discussed above it may be possible to design a functional assay for the production of colored naphthalene derivatives, such as 1,2-naphthalenediols, to screen for this second dehydrogenase using *Rhodococcus* ST-1 as the host strain.

Modifying the Indene Bioconversion Network

It is essential to consider two methodologies commonly used to modify a pathway when investigating the pharmaceutical application of this indene bioconversion system (Figure 4-4). One method is to

remove all competing and side reactions from the system. Genetically this can be done by creating null mutants using homologous recombination or amplifying the expression of other genes. This requires the cloning of most, if not all, of the genes involved in the system. The second method is to introduce the appropriate genes into a heterologous host to reconstitute the portions of the pathway necessary for the application. This method requires the cloning of the genes encoding the enzymes that produce the desired product. *A priori* there is no reason to favor either approach.

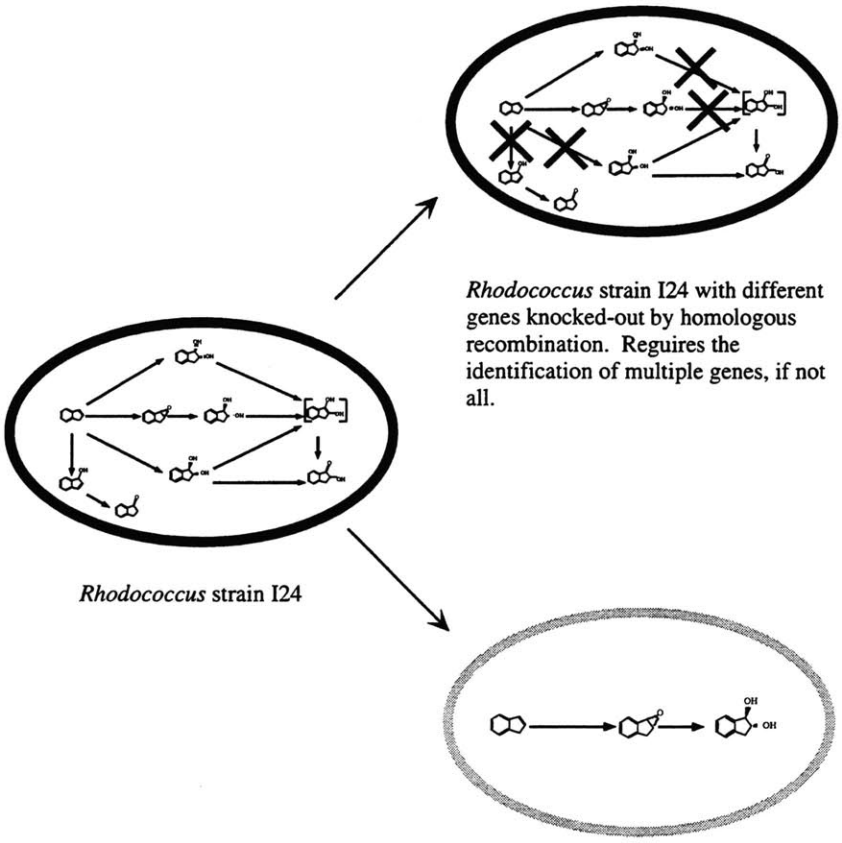


Figure 4-4: Modifying the Indene Bioconversion Pathway
There are two ways to optimize a pathway, by removing activities from the native system (upper right) or by expressing the desired activities in a naïve host (bottom right). Depending on the complexity of the system one methodology could be more direct than the other.

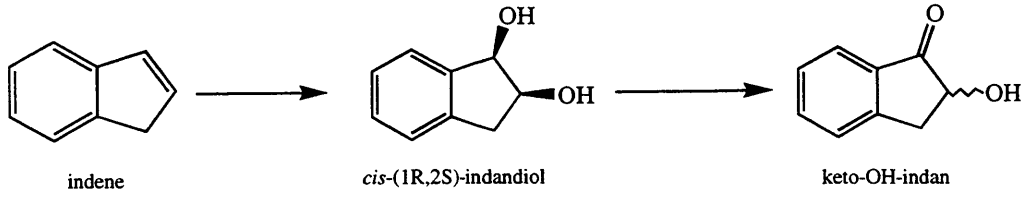
The indene bioconversion pathway can be modified by either methodology, the disruption method or the reconstitution method. The indene network has multiple redundant systems including oxygenases and dehydrogenases and the substrate range of these enzymes is unknown at this time. Thus modifying the indene bioconversion pathway via the removal method may be difficult because multiple genes need to be

cloned in order to remove their activities from the system. As seen with the *nidC* dehydrogenase null mutant, removing an activity does not always result in the predicted alteration. However, given the tools and genes identified there are two ways to modify the indene bioconversion pathway that may yield a useful modified strain. The first modification would be to remove the genes encoding the *nidAB* dioxygenase. This dioxygenase is responsible for production of at least some of the *cis*-(1R,2S)-indandiol in *Rhodococcus* strain I24. Removing the *nidAB* genes might eliminate production of the *cis*-(1R,2S)-indandiol. Another way to reduce or remove the unwanted indandiol enantiomer would be to overexpress the *nidC* diol dehydrogenase since it is specific for the *cis*-(1R,2S)-indandiol. These methods alone, or possibly combined, could produce a modified version of *Rhodococcus* strain I24 that produces only the other two indandiols.

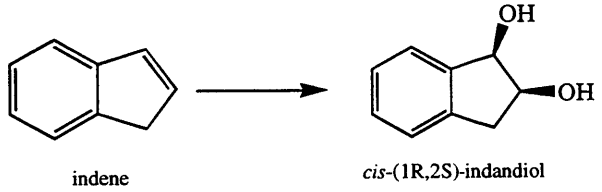
Reconstruction of part of the indene bioconversion pathway can result in a simplified system that may be useful for the production of an indandiol precursor molecule. Using the cosmid clone pR4 and the subclone pR4-10 described in Chapter Three, we demonstrated that the reconstruction methodology is viable for introducing a small portion of the indene pathway into a naïve host. For a summary of these simplified indene bioconversions see Figure 4-5. Due to the complexity of the indene bioconversion pathway and the potential for hidden activities, reconstitution may be a better route to an engineered strain in this case. Reconstruction requires the cloning of the other oxygenase systems in *Rhodococcus* strain I24. This could be attained in a variety of ways. Using the sequences available in the public databases and the sequence of the dioxygenase NidAB, homology based methods could be used to identify new genes by screening the cosmid library. The clustering of genes (e.g. the dioxygenase and the dehydrogenase) permits the use of both the dioxygenase and dehydrogenase as probes.

In addition to homology based methods, other functional screens could be developed and used to isolate genes encoding parts of the indene bioconversion pathway. For instance, mutagenized cells of *Rhodococcus* strain I24 or *Rhodococcus* ST-1 could be grown in the presence of naphthalene or toluene, and screened for mutants specifically in the oxygenase systems using nitro blue tetrazolium (NBT)/2,3,5-triphenyl-2H-tetrazloium chloride (TTC) system described in the introduction. Mutants in the oxygenase system should be different from the wild-type cells grown under the same conditions. Furthermore, if the naphthalene breakdown profile was determined it may be possible to use a naphthalene derivative (many of which are colored) in a functional screen to clone other genes. Similar analysis could be conducted with toluene as the substrate. The genes would ultimately be cloned by

A. *R. erythropolis* SQ1(pR4)



B. *R. erythropolis* SQ1(pR4-10)



C. *Rhodococcus* strain I24

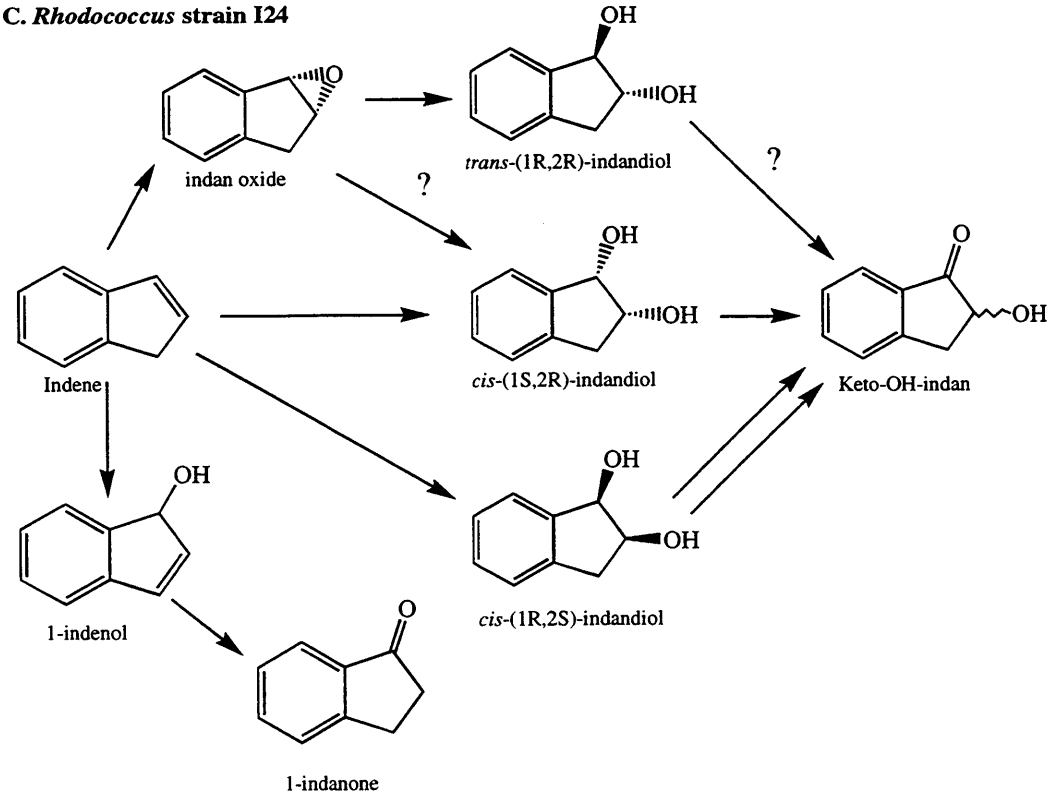


Figure 4-5: Summary of indene bioconversions
There are three strains that were examined for their ability to bioconvert indene. This diagram shows a summary of those bioconversions. A). *R. erythropolis* SQ1(pR4), the original cosmid clone, B) *R. erythropolis* SQ1(pR4-10), a subclone of pR4-10, C) *Rhodococcus* strain I24, a modified model.

complementation using the currently available cosmid library. To specifically target the monooxygenase activity it may be possibly to identify a substrate that can be converted to a colored epoxide that could be visualized. The identification of genes encoding the other oxygenase systems is essential for future modifications to *Rhodococcus* strain I24.

Given the model proposed in Figure 4-3, it may be imperative to clone the second dehydrogenase activity in order to reconstruct a strain for the production of an indandiol precursor. Although either *trans*-(1R,2R)-indandiol or *cis*-(1S,2R)-indandiol can be used as a precursor, a mixture of the two would need to be separated into its constituent parts. Consequently, it would be ideal to reconstruct a strain that produces only one of the appropriate indandiols. If the indan oxide is spontaneously breaking down into these two indandiols then a dehydrogenase may be employed to remove one enantiomer. We believe the strain *Rhodococcus* strain I24 already contains a useful dehydrogenase to accomplish this (see above). As previously discussed there a variety of ways to clone other dehydrogenases.

Conclusion

An increasing amount of research is being directed at the analysis of *Rhodococcus* species and their diverse metabolic capabilities. Rhodococci have the potential to play a major role in bioremediation and pharmaceutical synthesis. Researchers have only begun to realize the full potential of this species. Even though Rhodococci are diverse and not genetically well characterized, it is possible to genetically manipulate them. As more strains are characterized genetic analysis should become easier as commonalties between strains are determined.

The genetic analysis of *Rhodococcus* strain I24 has been a rewarding research experience. Basic microbiology, molecular biology and biochemistry were combined towards the analysis of the strain and its ability to carry out the bioconversion of indene. The *Rhodococcus* strain I24 indene bioconversion pathway is a complex network. From the analyses presented in this work we have shown it is possible to develop a framework in which to analyze newly identified strains of bacteria and their unique attributes. We established a model pathway for the bioconversion of indene and developed methods to identify genes that encode the enzymes in that pathway. In order to analyze the genes and the enzymes encoded by them, we developed genetic tools to modify the strain and enzymatic assays to investigate the capabilities of the enzymes. The model for the indene bioconversion pathway was modified based on the genetic and enzymatic analysis conducted using these tools, providing new insights for future analysis of the system. The framework, the genetic tools and the biochemical tools/assays developed in this work will be invaluable for continuing investigations of the indene bioconversion pathway in *Rhodococcus* strain I24 and the development of a genetically engineered strain to produce a precursor for an important therapeutic, the HIV-1 protease inhibitor Crixivan™.

References

Allen CCR, Boyd DR, Dalton H, Sharma ND, Brannigan I, Kerley NA, Sheldrake GN, Taylor SC (1995) Enantioselective bacterial biotransformation routes to *cis*-diol metabolites of monosubstituted benzenes, naphthalene and benzocycloalkenes of either absolute configuration. J. Chem. Soc. Chem. Commun. 1995: 117-118

Allen CCR, Boyd DR, Larkin MJ, Reid KA, Sharma ND, Wilson K (1997) Metabolism of Napthalene, 1-Napthol, Indene, and Indole by *Rhodococcus* sp. Strain NCIMB 12038. Appl Environ Microbiol 63: 151-155

Chartrain M, Jackey B, Taylor C, Sandford V, Gbewonyo K, Lister L, DiMichele L, Hirsch C, Heimbuch B, Maxwell C, Pascoe D, Buckland B, Greasham R (1998) Bioconversion of indene to *cis*-(1S,2R)-indandiol and *trans*-(1R,2R)-indandiol by *Rhodococcus* species. J Ferment Bioeng 86: 550-558

Dabrock B, Kesseler M, Averhoff B, Gottschalk G (1994) Identification and characterization of a transmissible linear plasmid from *Rhodococcus erythropolis* BD2 that encodes isopropylbenzene and trichloroethene catabolism. Appl Environ Microbiol 60: 853-60

Ensley BD, Gibson DT (1983) Naphthalene dioxygenase: purification and properties of a terminal oxygenase component. J Bacteriol 155: 505-11

Gibson DT, Resnick SM, Lee K, Brand JM, Torok DS, Wackett LP, Schocken MJ, Haigler BE (1995) Desaturation, Dioxygenation, and Monooxygenation Reactions Catalyzed by Napthalene Dioxygenase from *Pseudomonas* sp. Strain 9816-4. J Bacteriol 177: 2615-2621

Haigler BE, Gibson DT (1990a) Purification and properties of ferredoxinNAP, a component of naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816. J Bacteriol 172: 465-8

Haigler BE, Gibson DT (1990b) Purification and properties of NADH-ferredoxinNAP reductase, a component of naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816. J Bacteriol 172: 457-64

Kauppi B, Lee K, Carredano E, Parales RE, Gibson DT, Eklund H, Ramaswamy S (1998) Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2- dioxygenase. Structure 6: 571-86

Masai E, Sugiyama K, Iwashita N, Shimizu S, Hauschild JE, Hatta T, Kimbara K, Yano K, Fukuda M (1997) The *bphDEF* meta-cleavage pathway genes involved in biphenyl/polychlorinated biphenyl degradation are located on a linear plasmid and separated from the initial *bphACB* genes in *Rhodococcus* sp. strain RHA1. Gene 187: 141-149

Nakazawa T, Yokota T (1977) Isolation of a mutant TOL plasmid with increased activity and transmissibility from *Pseudomonas putida (arvilla)* mt-2. J Bacteriol 129: 39-46

Parales RE, Parales JV, Gibson DT (1999) Aspartate 205 in the catalytic domain of naphthalene dioxygenase is essential for activity. J Bacteriol 181: 1831-1837

Resnick SM, Lee K, Gibson DT (1996) Diverse reactions catalyzed by naphthalene dioxygenase from *Pseudomonas* sp strain NCIB 9816. J Ind Micro 17: 438-57

Shields MS, Reagin MJ, Gerger RR, Campbell R, Somerville C (1995) TOM, a new aromatic degradative plasmid from *Burkholderia (Pseudomonas) cepacia* G4. Appl Environ Microbiol 61: 1352-6

Subramanian V, Liu TN, Yeh WK, Serdar CM, Wackett LP, Gibson DT (1985) Purification and properties of ferredoxinTOL. A component of toluene dioxygenase from *Pseudomonas putida* F1. J Biol Chem 260: 2355-63

Wackett LP, Kwart LD, Gibson DT (1988) Benzylic monooxygenation catalyzed by toluene dioxygenase from *Pseudomonas putida*. Biochemistry 27: 1360-7

Williams PA, Worsey MJ (1976) Ubiquity of plasmids in coding for toluene and xylene metabolism in soil bacteria: evidence for the existence of new TOL plasmids. J Bacteriol 125: 818-28

Yen KM, Gunsalus IC (1982) Plasmid gene organization: naphthalene/salicylate oxidation. Proc Natl Acad Sci U S A 79: 874-8

Appendix A: Construction of a *nidAB* Null Mutant

In order to analyze the indene bioconversion network construction of a null mutant of the *nidAB* dioxygenase was attempted. A knock-out construct, pST124, was made. Most of the *nidAB* open reading frames was replaced with the gene encoding spectinomycin antibiotic resistance, *aadA*, from the omega interposon (Prentki et al. 1991). This plasmid has the RP4 mobility element and this construct can be mobilized from *E. coli* S17-1 to *Rhodococcus* strain I24. A standard filter mating was conducted to introduce the plasmid into *Rhodococcus* strain I24. The protocol is described in Chapter Three. Transconjugants were selected on LB plates supplemented with spectinomycin (200 μg/ml) and nalidixic acid (10 μg/ml). The plates were incubated at 30°C and colonies appeared after about one week. This was about the same amount of time needed in other conjugation experiments with *Rhodococcus* strain I24. Approximately 15 candidate strains were obtained from several mating experiments.

Genomic DNA of the candidate strains was analyzed by PCR and by southern blotting. The control PCR reactions in the PCR analysis were unreliable. Consequently PCR analysis was not used to determine if any of the candidate strains were *nidAB* null mutants. The primers may not have been optimal. The southern blot analysis was more successful, but required optimization. The best southern blot results were obtained when the DNA was transferred in 20X SSC rather than in NaOH. The hybridization temperature used in these experiments was 42°C. Recent work has suggested that the hybridization can be further improved if incubated at 65°C (A. Dexter, unpublished results). A variety of different probes including a portion of the *nidA* gene, the *nidC* gene, and the *aadA* gene, were used in this analysis. None the spectinomycin resistant strains represented a deletion of the *nidAB* genes. Furthermore, none of these strains carried a gene homologous to the *aadA* gene. This suggests that *Rhodococcus* strain I24 can become resistant to spectinomycin relatively easily, although a rate was never calculated.

In order to repeat this experiment it would be necessary to reconstruct the knock-out construct, replacing the *aadA* spec gene with a different antibiotic marker. The thiostrepton (*tsr*) gene is the antibiotic marker recommended as it was used successfully in the deletion of the *nidC* gene. I originally chose the spec gene as a marker in case a double mutant lacking the dioxygenase and the dehydrogenase, a Δ*nidABC* strain, was needed. Given the fact that *nidABC* are probably in an operon, it may be unnecessary to prepare a double mutant, as disruption of the dioxygenase subunits may result in the disruption of the dehydrogenase gene as well. Hence, using the *tsr* should not be a problem for constructing a null mutant of *nidAB.*

Prentki P, Binda A, Epstein A (1991) Plasmid vectors for selecting IS1-promoted deletions in cloned DNA:Sequence analysis of the Omega interposon. Gene 103:17-23

Appendix B: Diol Dehydrogenase Enzyme Assay - The Numbers

Although in crude lysates the activity rates are not very meaningful for comparison to other dehydrogenases, these numbers may be helpful toward the purification of this enzyme. Below are tables with the rates in AU/min as derived from the spectrophotometer. These rates were determined by using the initial slope of the reaction. Total reaction time was 600 sec. However, the reaction was complete within the first 200 seconds. The enzyme assay conditions are described in Chapter Three.

The optical density of the culture prior to lysate production significantly reduces the enzyme activity. Cultures with higher OD_{600} (4.0-5.0) generally had less activity than those with an OD_{600} of 2.0-3.5. This was deceiving at first because protein concentrations of the higher OD600 cultures were similar to the protein concentrations of lysates derived from cultures with a lower optical density. It is possible that the higher OD is the consequence of dead cells, which can contribute to the protein concentration, but that the dehydrogenase enzyme is unstable under these conditions.

All enzyme assays in the table were conducted with *cis*-(1R,2S)-dihydro-naphthalenediol as the substrate. Rates are determined immediately after lysate production. The lysate still contains activity the next day, but the rate is reduced by about half.

R. erythropolis SQ1(pST117) Lysate		OD_{600}	Total Protein (μg/ul)	Rate (AU/min)
Number	Date			
1	6-8-98	2.76	30.58	0.168
2	6-8-98	2.52	34.36	0.125
3	6-15-98	2.6	16.02	0.304
4	6-15-98	2.18	11.95	0.298
5	6-15-98	2.18	19.29	0.235
				0.273
				0.268
6	6-18-98	2.48	12.9	0.229
7	6-18-98	2.88	14.9	0.199
8	9-14-98	2.28	8.9	0.732
9	9-14-98	2.48	11.3	0.298
10	9-14-98	3.34	10.9	0.333
11	9-23-98	4.68	9.97	0.089
12	9-23-98	3.7	6.22	0.062
13	9-23-98	4.94	10.66	0.112

Appendix C: Indandiol Oxidation – Another Way to Present the Data

The analysis of indandiol oxidation by *R. erythropolis* SQ1 derivative strains, *Rhodococcus* I24, and *Rhodococcus* ST-1 can be presented in tabular fashion. Those tables are presented below.

Indandiol Oxidation of *Rhodococcus* I24 and *Rhodococcus* ST-1 (*Rhodococcus* I24Δ*nidC*)
See Chapter Three

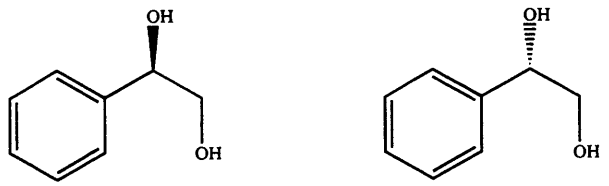
Strain	*cis*-(1R,2S)-indandiol			*cis*-(1S,2R)-indandiol			
	At 0 hrs	At 24 hrs		At 0 hrs	At 24 hrs		
	[initial] mg/L	[*cis*] mg/L	[keto] mg/l	[initial] mg/L	[*cis*] mg/L	[keto] mg/l	
Rhodococcus I24 Δ*nidC*	18.70	12.90	2.35	53.36	50.04	5.48	
	16.20	13.60	2.80	36.73	28.19	3.74	
	40.65	29.30	6.10				
	25.18	18.60	3.75	45.05	39.12	4.61	Average
Rhodococcus I24	43.90	6.10	23.90	43.20	37.60	4.93	
	19.12	5.01	9.60	32.73	31.87	5.14	
	15.31	4.53	7.38				
	26.11	5.21	13.63	37.97	34.74	5.04	Average

***In vivo* Indandiol Oxidation**
See Chapter Three for method and an alternative presentation of the data.

Strains	Genes Expressed	*cis*-(1R,2S)-indandiol			*cis*-(1S,2R)-indandiol			*trans*-(1R2R)-indandiol			
		at 0 hrs	at 24 hrs		at 0 hrs	at 24 hrs		at 0 hrs	at 24 hrs		
		[initial] mg/L	[*cis*] mg/L	[keto] mg/l	[initial] mg/L	[*cis*] mg/L	[keto] mg/l	[initial] mg/L	[*cis*] mg/L	[keto] mg/l	
R. erythropolis SQ1 (pAPE12)	none	6.99	8.10	0.00	42.57	40.97	0.00	24.20	25.20	0.00	
		7.69	7.09	0.00				31.80	31.80	0.00	
		7.34	7.60	0.00	42.57	40.97	0.00	28.00	28.50	0.00	Average
R. erythropolis SQ1 (pR4)	Dioxygenase Dehdyrogenase	29.25	15.01	47.40	41.33	35.22	0.00	26.69	23.10	0.00	
		9.69	0.00	18.29	59.58	58.10	0.00	49.50	41.20	0.00	
		19.47	7.51	32.85	50.46	46.66	0.00	38.10	32.15	0.00	Average
R. erythropolis SQ1 (pR4-10)	Dioxygenase	28.24	30.26	0.00	33.65	41.50	0.00	27.72	26.09	0.00	
		26.16	24.60	0.00	41.90	49.00	0.00	78.90	43.50	0.00	
		6.97	6.36	0.00							
		20.46	20.41	0.00	37.78	45.25	0.00	53.31	34.80	0.00	Average

Appendix D: Dehydrogenase Enzyme Assay - The Styrene Glycol Story

The analysis of the dehydrogenase substrate specificity led to testing of alternative substrates. A compound similar to the indandiols and dihydro-naphthalenediol is styrene glycol.



R-Styrene Glycol S-Styrene Glycol

Figure C-1: Structure of Styrene Glycol
The structures of the two enantiomers of styrene glycol are shown.

Dehydrogenase enzyme assays, as described in Chapter Three, using both enantiomers of styrene glycol as substrates were conducted. In addition, different buffer systems were analyzed for their effects on the dehydrogenase assay. The strains analyzed include *R. erythropolis* SQ1(pST117) expressing the NidC dehydrogenase, and the control strain *R. erythropolis* SQ1(pAPE12) containing an empty expression vector. Each assay was repeated with three lysates. The results are shown in the table below. The pH of the buffer had an effect on substrate range. That is, the R-styrene glycol could be used as a substrate with a reaction buffer with a pH 9. Analysis of the control lysate, *R. erythropolis* SQ1(pAPE12), demonstrates that dehydrogenase activity that can use R-styrene glycol as a substrate is derived from the host strain *R. erythropolis* SQ1, not from the NidC enzyme encoded on the plasmid pST117. Furthermore, the *R. erythropolis* SQ1 dehydrogenase could be stereospecific as it is unable to use S-styrene glycol as a substrate at pH 9. Further investigation is required.

Compound	100 mM K^+ Phosphate Buffer pH 7.6		100 mM K^+ Phosphate Buffer pH 9.15		50 mM Tris-Propane, 0.8 M NaCl pH 9.5	
	pST117	pAPE12	pST117	pAPE12	pST117	pAPE12
cis-(1R,2S)-dihydro-naphthalenediol	+	-	+	-	nd[a]	nd
cis-(1R,2S)-indandiol	+	-	nd	nd	nd	nd
cis-(1S,2R)-indandiol	-	-	nd	nd	nd	nd
R-styrene glycol	-[b]	nd	+	+	+	nd
S-styrene glycol	-[b]	nd	-	nd	nd	nd

[a] nd, not determined
[b] assays conducted with and without NaCl

Appendix E: Plasmid Instability

Rearrangement of plasmid DNA, especially large plasmids, is not uncommon. Plasmids and cosmids containing DNA derived from *Rhodococcus* strain I24 exhibit some rearrangement capabilities. Rearrangement was first encountered in the analysis of the cosmid clone pR4 and the subclone pR4-10. When these plasmids are transformed into *E. coli* and the cells are grown to saturation in a volume larger than 50 ml, the plasmids rearrange. Rearrangement was demonstrated by repeating the indigo formation assay as described in Chapter Two. The resulting plates of *E. coli*(pR4) and *E. coli*(pR4-10) exhibited a mixture of blue and white colonies. When the plasmid DNA content of the blue colonies and the white colonies derived from *E. coli*(pR4-10) was analyzed by restriction enzyme digest, all of the blue colonies produced one restriction pattern and all of the white colonies produced a second pattern. The same results were obtained from the analysis of *E. coli*(pR4) derived colonies. Rearrangement of pR4 and pR4-10 in *E. coli* can be prevented by growing 5 ml cultures. All initial analysis of the cosmids clones and subclones was conducted with small cultures of this type which explains why in the initial analysis of the cosmid clone and subclone that rearrangement was not detected at that time. Consequently, all large scale DNA preps of these plasmids requires the pooling of multiple small scale preparations of DNA in order to obtain a unique plasmid population.

Cosmid clones expressed in the host strain *Rhodococcus erythropolis* SQ1 also exhibit rearrangement. This has not been seen with cosmid pR4 or pR4-10, but rather with other cosmid clones from the *Rhodococcus* strain I24 cosmid library. Some cosmids when transformed into *R. erythropolis* SQ1 undergo rearrangement. Once again the rearrangement was identified during analyses of plasmid DNA extracted from the cells. The plasmid DNA was compared to plasmid DNA derived from *E. coli* via restriction enzyme digest analysis. No pattern of rearrangement has been determined. Furthermore, in a population of transformants rarely will every transformant analyzed produce rearranged plasmids.

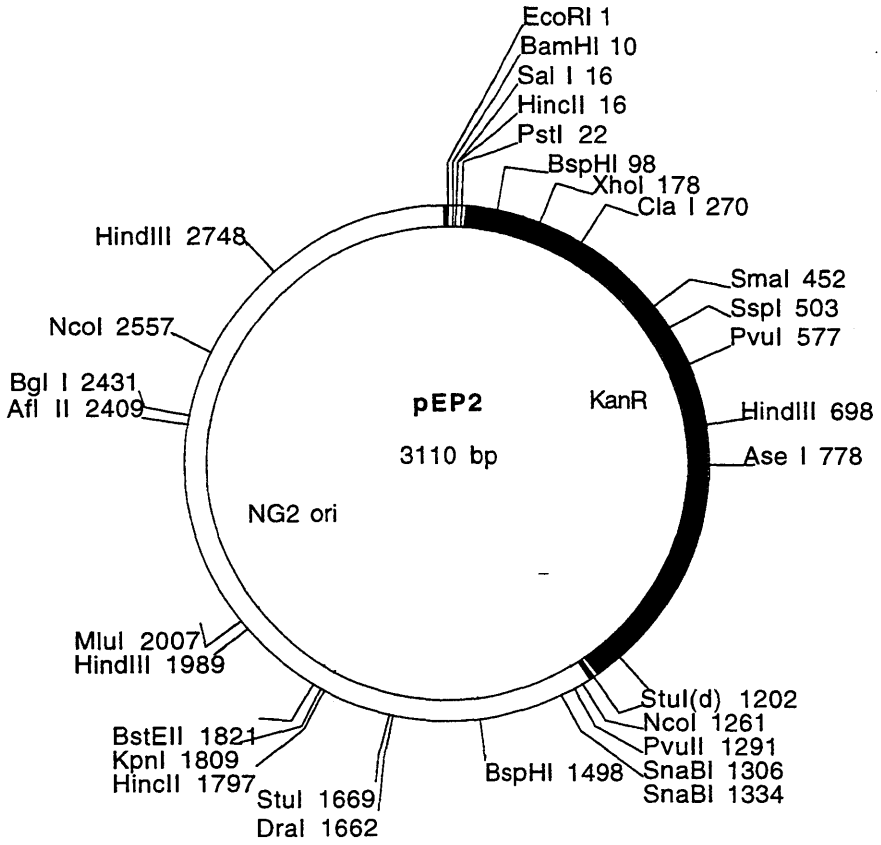
Conjugation experiments with *E. coli* S17-1 to *Rhodococcus* strain I24 demonstrate that rearrangement of plasmids can also occur in *Rhodococcus* strain I24 (P. Lessard and X. Shaver, unpublished results). Approximately one out of every three transconjugants is rearranged. Interestingly, the plasmid in this case is partially derived from another strain of *Rhodococcus*, *Rhodococcus* strain B264-1.

It is unclear what is responsible for inducing plasmid instability. The G+C content of Rhodococci is fairly high, approximately 65%. This could lead to regions that are highly similar that could undergo spontaneous recombination at a certain frequency in a heterologous host. However, sequence analysis of pR4-10 did not reveal any repeat regions that could contribute in this manner. It is also possible that all of the strains used in these analyses (*E. coli* JM109 or DH5α, *R. erythropolis* SQ1, and *Rhodococcus* strain

I24) may contain a recomibinase or recombinase-like enzyme that recognizes something unique to Rhodococcal DNA inducing plasmid instability. Further analysis of this phenomenon could be useful.

Appendix E: Plasmid Maps

The following pages contain the maps for plasmids constructed for this work. This information is meant to be a permanent record to facilitate the work of others who may continue with this project.



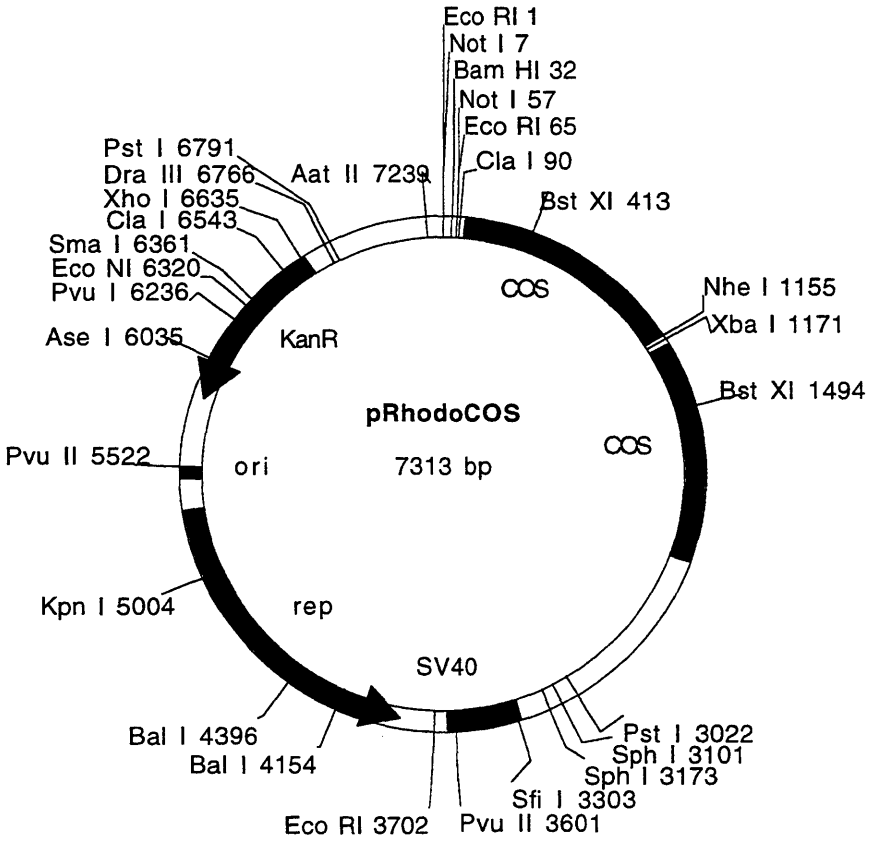
Plasmid name: pEP2

Plasmid size: 3110 bp

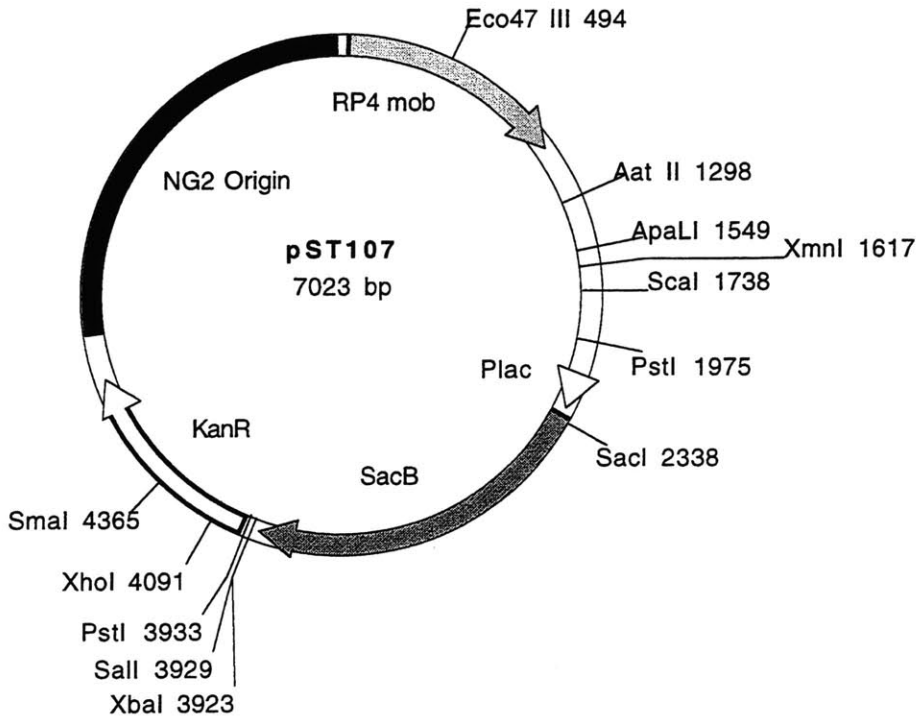
Constructed by: Adrian Hodgson, CSIRO

Construction date:

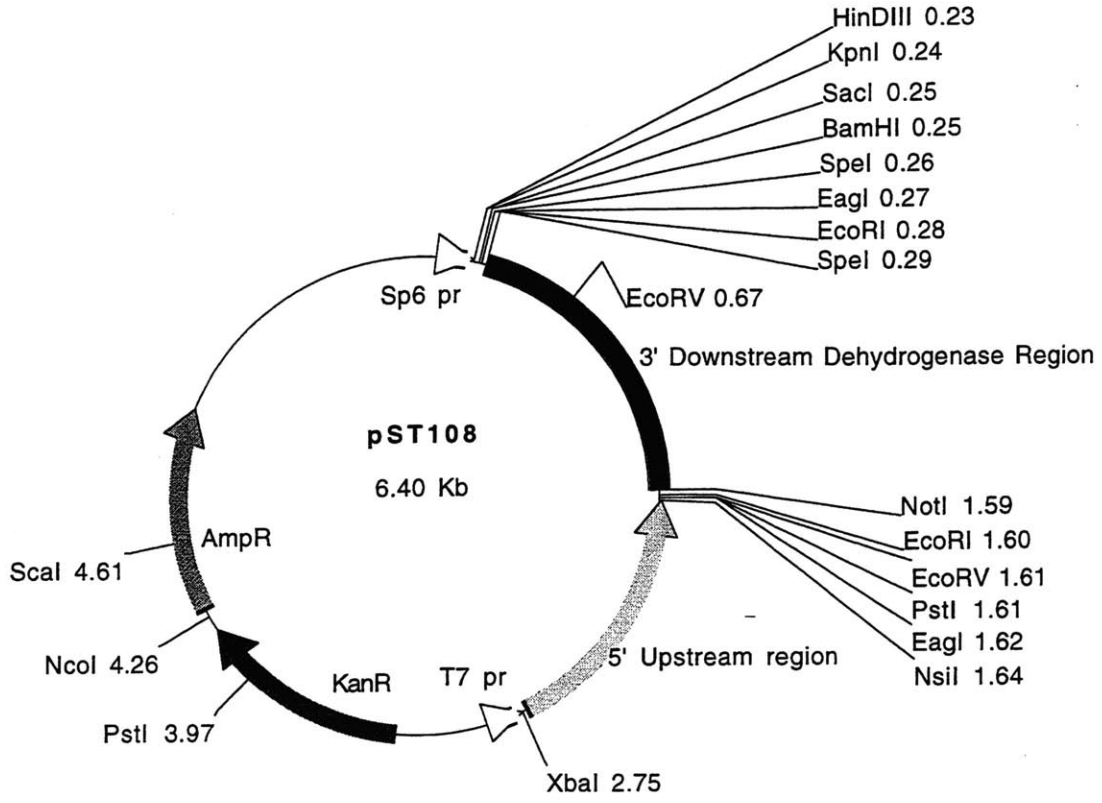
Comments/References: Received from A.J. Pittard, University of Melbourne; Kanamycin resistance marker derived from pUC4K; origin of replication derived from NG2; replicates well in E. coli, Corynebacterium and in Rhodococcus



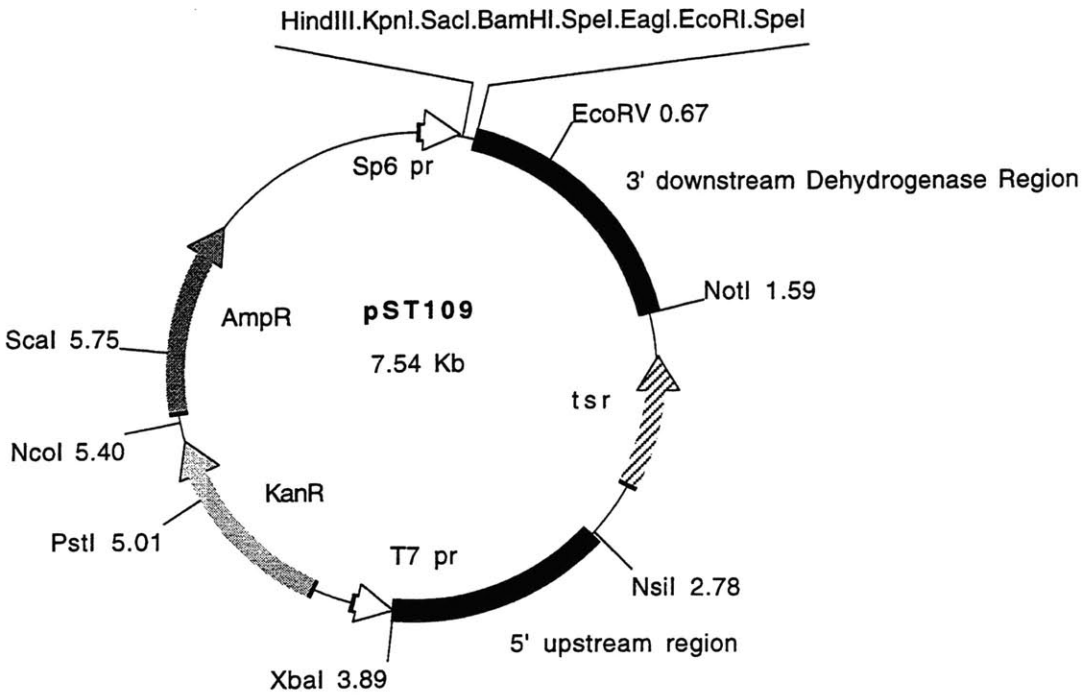
Plasmid name: pRhodoCOS
Plasmid size: 7313 bp
Constructed by: Sheri Treadway
Construction date: 6 August 1997
Comments/References: Constructed by replacing the origin and selectable marker in SuperCOS with those from pEP2; pEP2 ligated as a BamHI-SalI (blunted) fragment into the BglII-ScaI sites of SuperCOS; should replicate in both E. coli and Rhodococcus; KanR



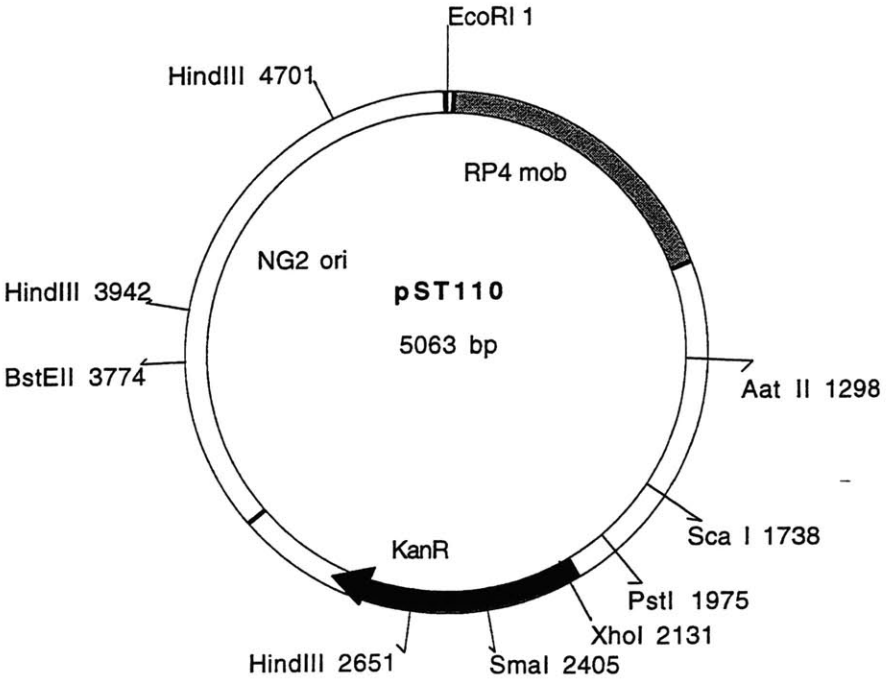
Plasmid name: pST107
Plasmid size: 7023 bp
Constructed by: S. Treadway
Construction date: Jan 1998
Comments/References: pAL226 cut with PstI ligated with a PstI fragment from pWS1 carrying the sacB gene under control of the lac promoter. Cloning sites: SalI, XbaI, ScaI, ApaLI, ApaI.



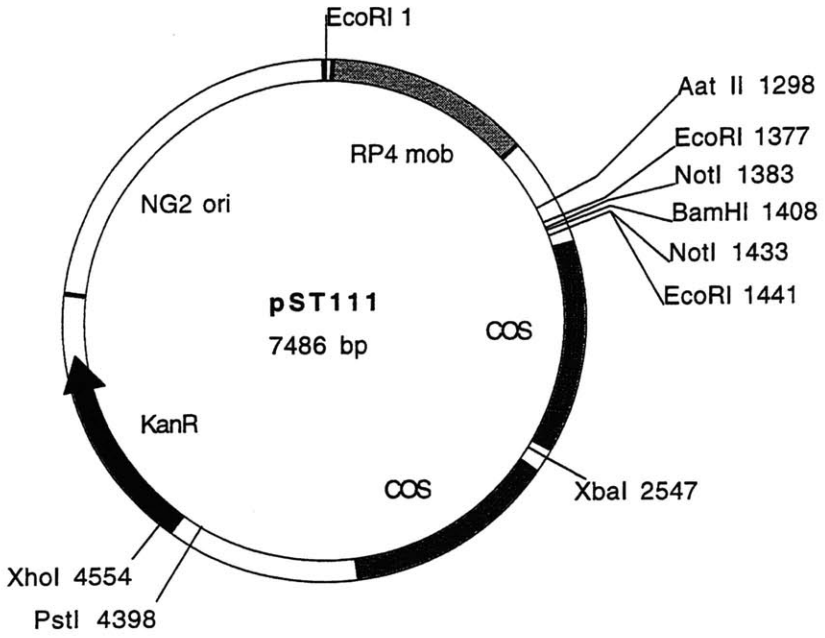
Plasmid name: pST108
Plasmid size: 6.40 kb
Constructed by: S.Treadway
Construction date: 3/12/98
Comments/References: pST114 cut with Xba and Nsi. Insert is an Xba-Nsi fragment from pST112. Sizes are approximate.



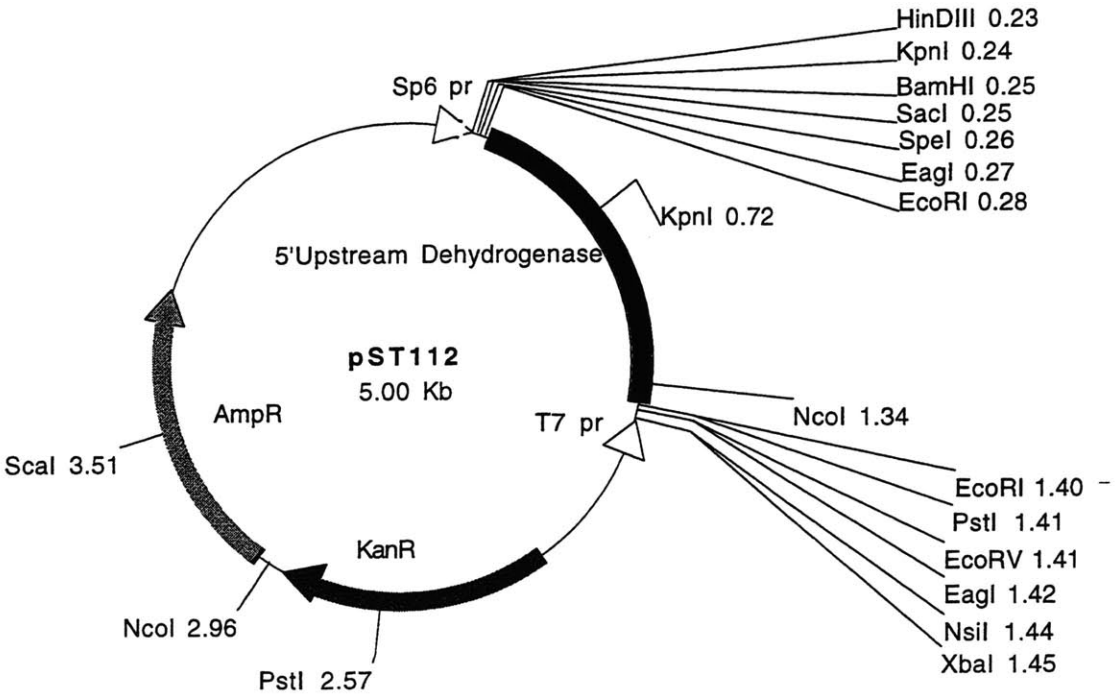
Plasmid name: pST109
Plasmid size: 7.54 kb
Constructed by: S. Treadway
Construction date: 3/18/98
Comments/References: pST108 cut with NotI and NsiI with a 1.14 Kb Bsp120I-PstI insert from pAL231 contianing the tsr gene (thiostrepton). Preliminary Dehydrogenase knock-out construct. Positions of antibiotic genes are appoximate.



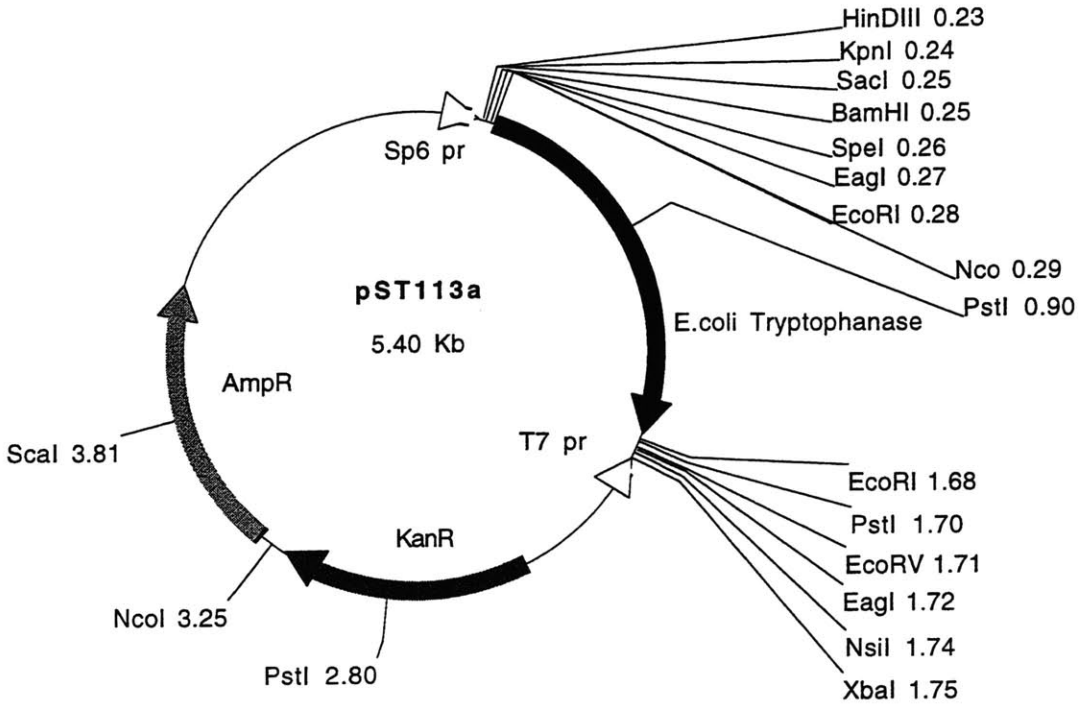
Plasmid name: pST110
Plasmid size: 5063 bp
Constructed by: S. Treadway
Construction date: Jan 1998
Comments/References: Derivative of pAL226 with the BamHI site removed using Klenow and filling in.



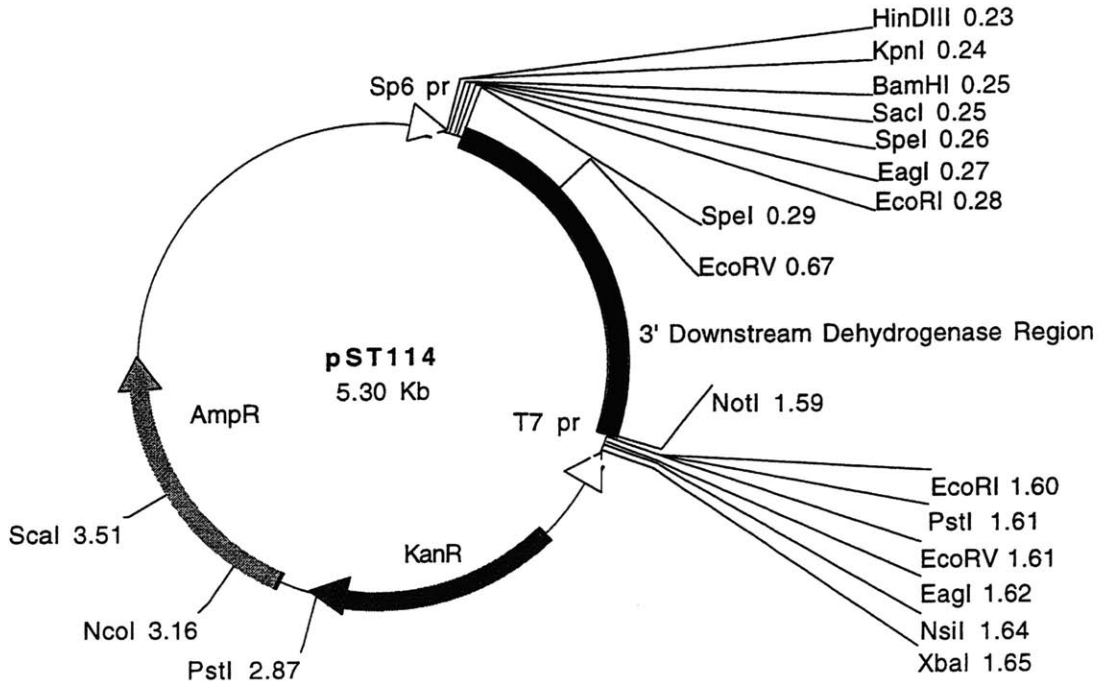
Plasmid name: pST111
Plasmid size: 7486 bp
Constructed by: S. Treadway
Construction date: Jan 1998
Comments/References: pST110 cut with AatII and PstI. Insert is a 3.2 kb AatII-PstI fragment from pRhodoCOS containing the COS sites and BamHI site for cosmid cloning.



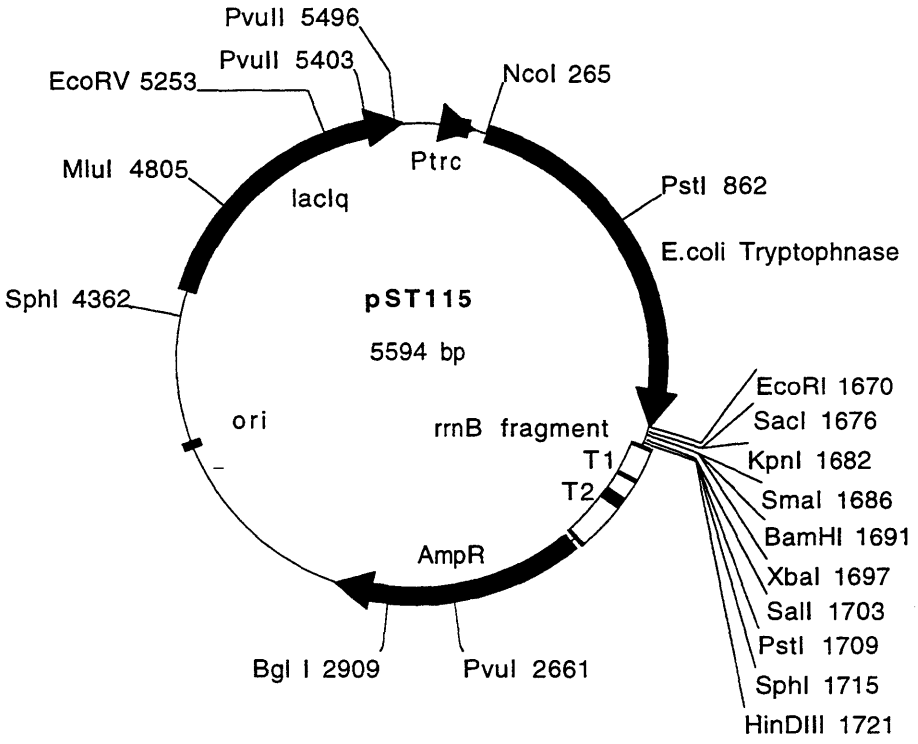
Plasmid name: pST112
Plasmid size: 5.00 kb
Constructed by: S.Treadway
Construction date: 2/14/98
Comments/References: pCR2.1-Topo (Invitrogen) with 1.1 kb PCR fragment of 5' Upstream Region from R4 dehydrogenase using primers DH-1 and DH-2.



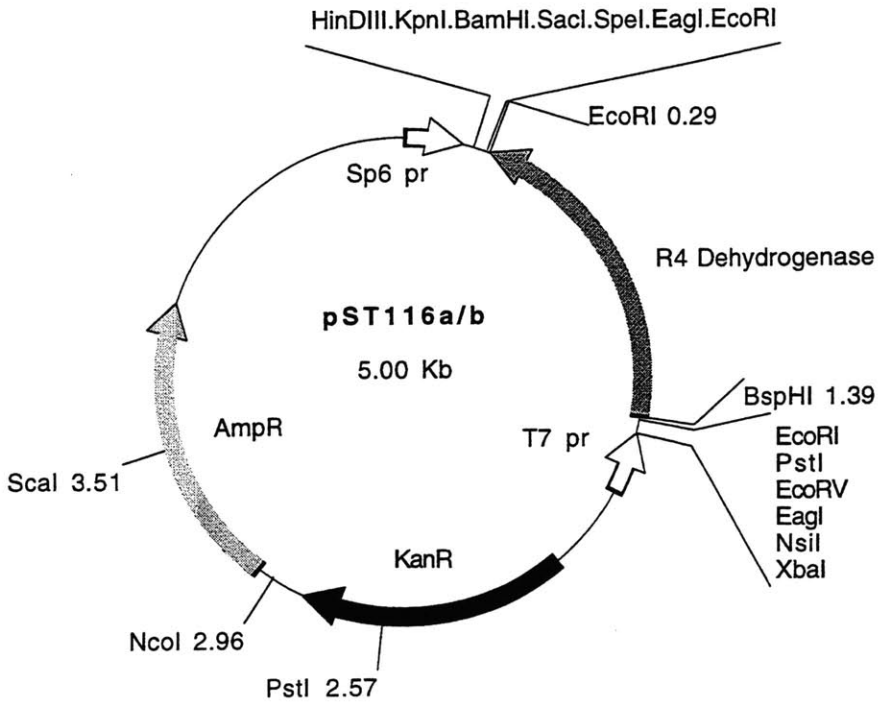
Plasmid name: pST113a
Plasmid size: 5.40 kb
Constructed by: S.Treadway
Construction date: 2/14/98
Comments/References: pCR2.1-Topo with 1.4 kb PCR fragment of E.coli Tryptophanse with primers 5'TrpNco and 3'Trp. There is a pST113b which has the trp gene in the opposite orientation. 113a Box 2 slot 78, 133b box 2 slot 79.



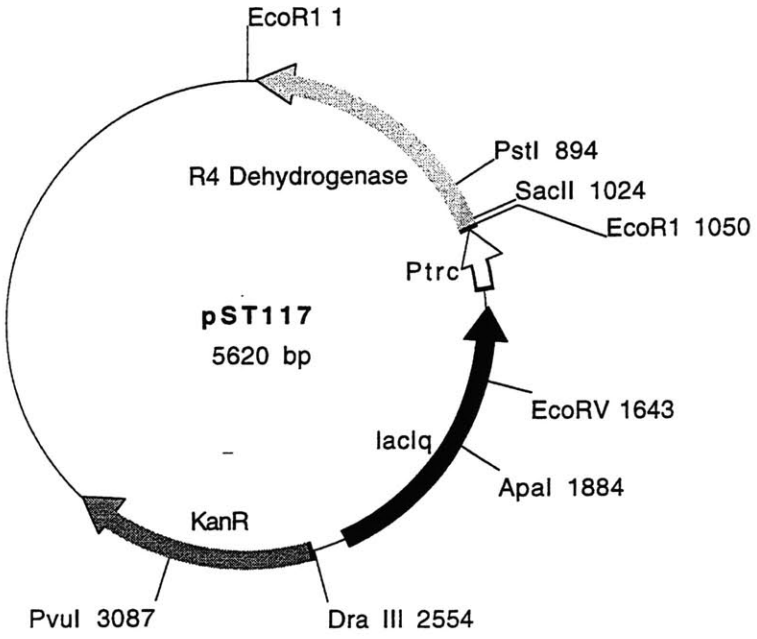
Plasmid name: pST114
Plasmid size: 5.30 kb
Constructed by: S.Treadway
Construction date: 2/23/98
Comments/References: pCR2.1-Topo with ~1.4 PCR fragment of 3' Downstream Region from R4 dehydrogenase using primers DH-3 and DH-4. Site distances are approximate.



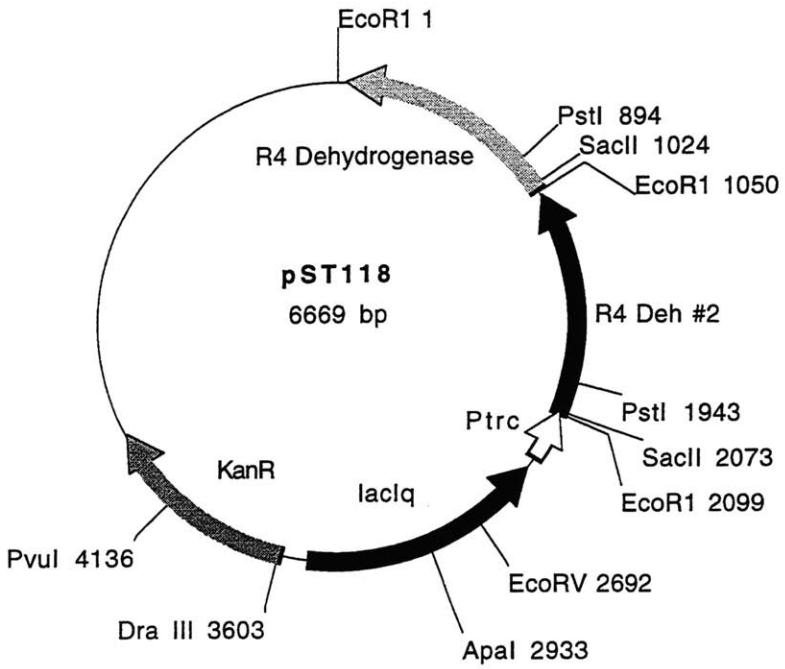
Plasmid name: pST115
Plasmid size: 5594 bp
Constructed by: S. Treadway
Construction date: 3/6/98
Comments/References: E.coli Tryptophanase moved as an Nco I/EcoRi fragment from pST113a, into the NcoI and EcoRI sites of pTRC99a.



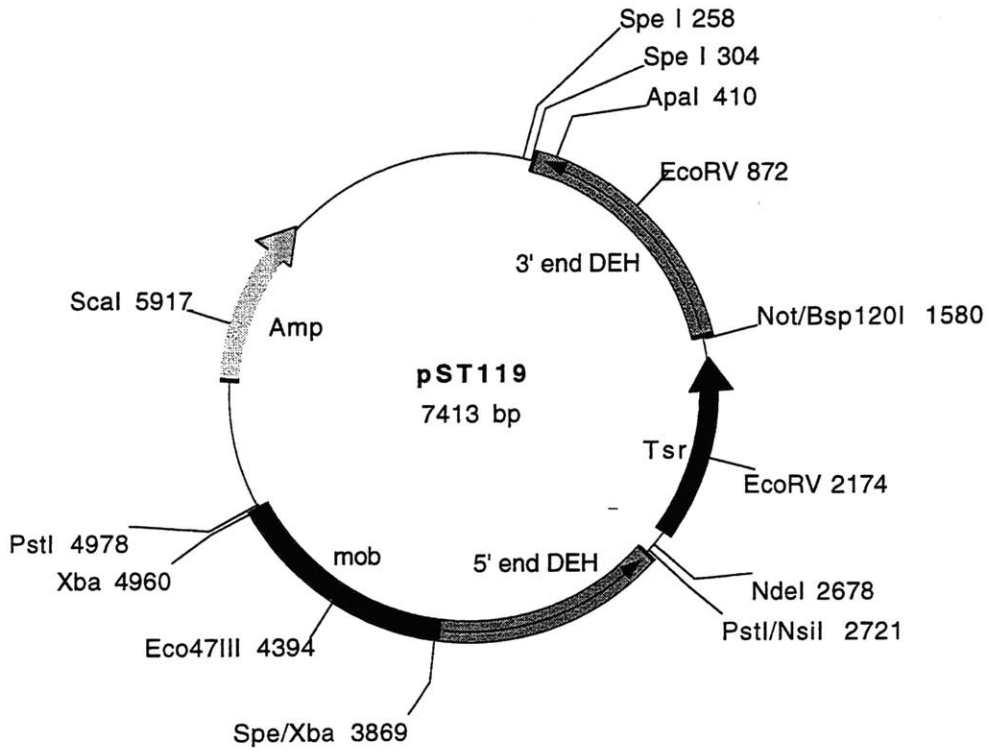
Plasmid name: pST116a/b
Plasmid size: 5.00 kb
Constructed by: S. Treadway
Construction date: 2/14/98
Comments/References: pCR2.1-Topo with 1.1 kb PCR fragment of the full length R4 dehydrogenase using primers 5'-dehydro and 3'-dehydro. pST116b has the insert in the opposite orientation of that shown above.



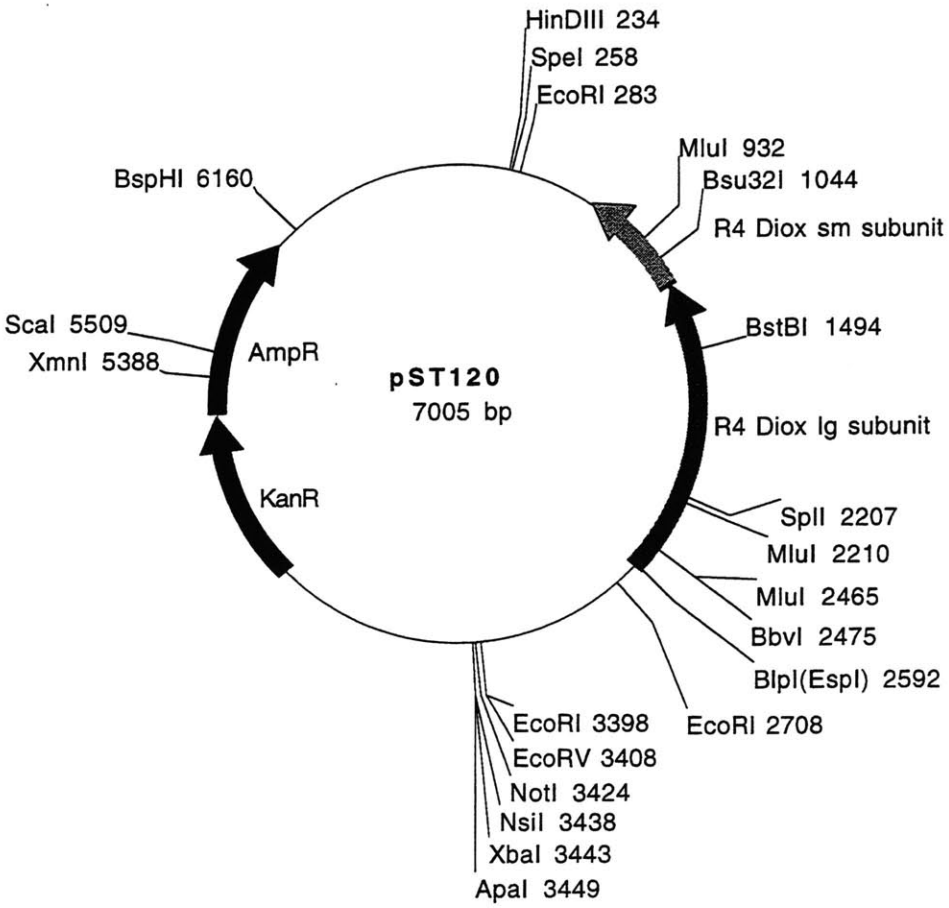
Plasmid name: pST117
Plasmid size: 5620 bp
Constructed by: S. Treadway
Construction date: 4/8/98
Comments/References: pAPE12 cut with EcoR1 and SAP treated. Insert from pST116a cut with EcoR1. NOTE: pAPE12 was made by Avi Rhodal by cutting pEP2 with PstI and SalI, inserting Ptrc as a Nsi-Sal fragment from pTRC99a.



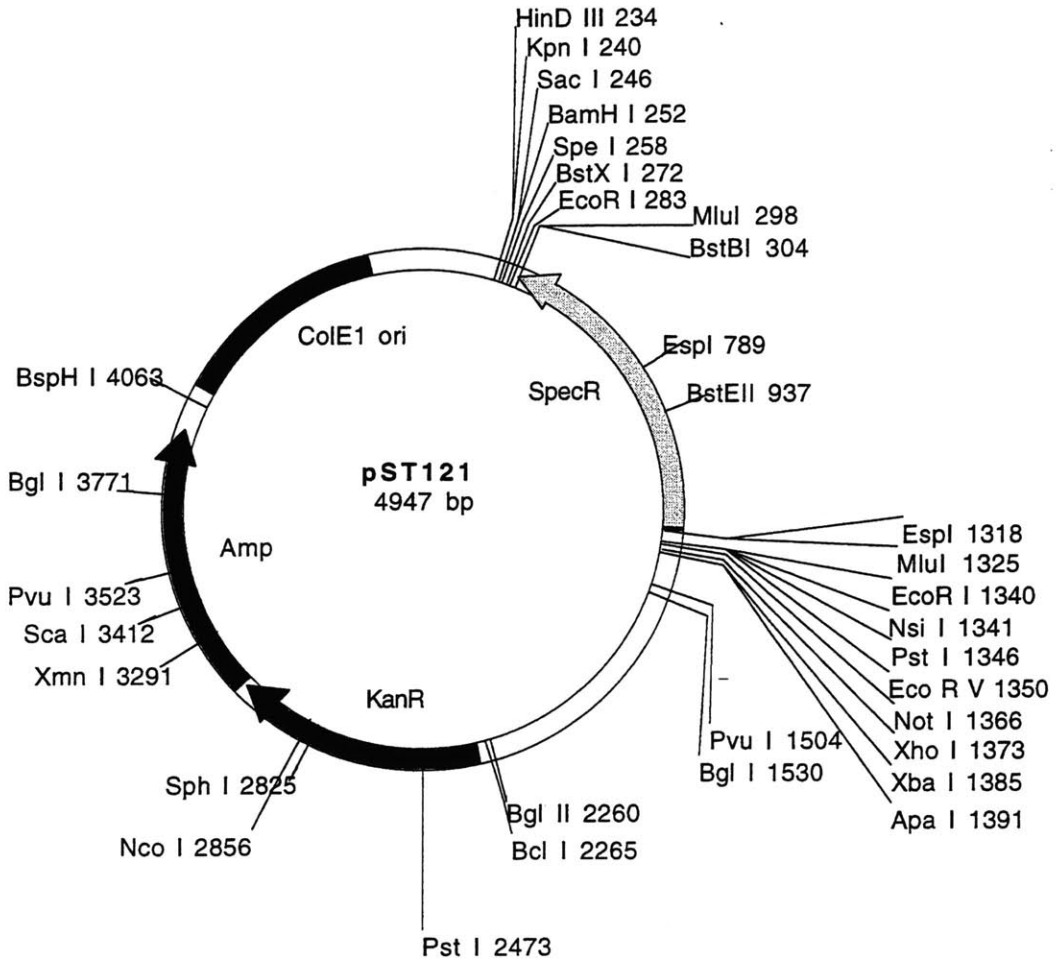
Plasmid name: pST118
Plasmid size: 6669 bp
Constructed by: S. Treadway
Construction date: 4/8/98
Comments/References: pAPE12 cut with EcoR1 and SAP treated. Insert from pST116a cut with EcoR1. This plasmid has a double insert, in tandem.



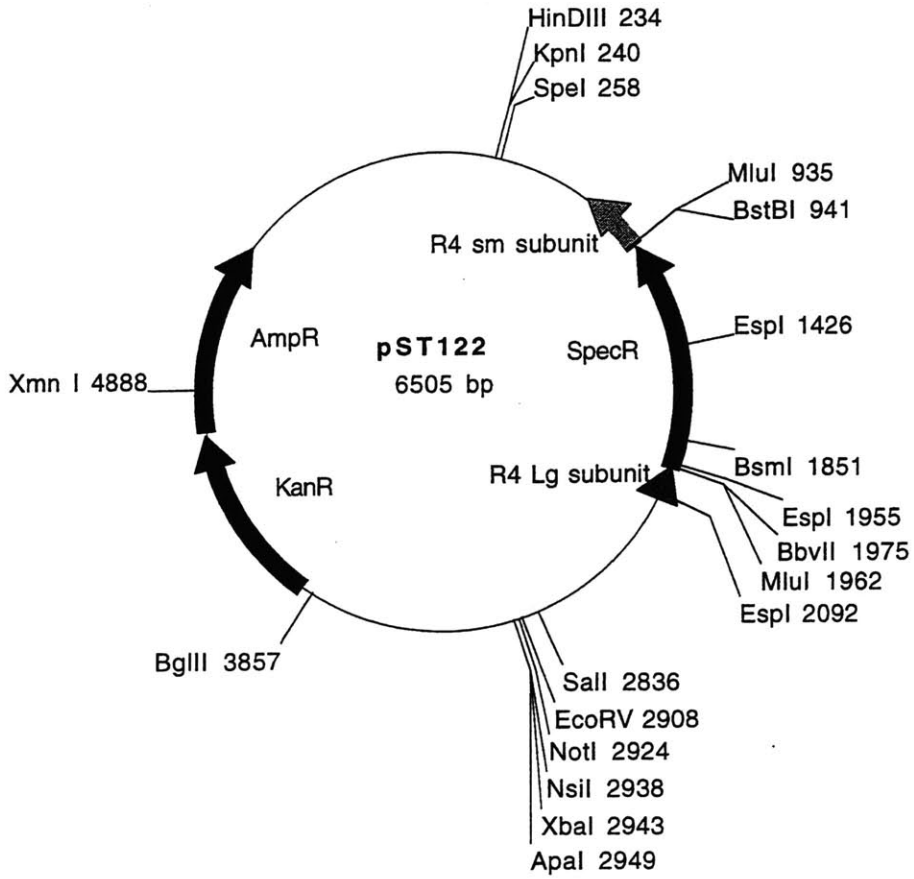
Plasmid name: pST119
Plasmid size: 7413 bp
Constructed by: S. Treadway
Construction date: April 1998
Comments/References: pST109 cut with PstI and XbaI. Insert is a PstI-SpeI fragment from pAL235 containing the mobility element RP4. Plasmid contains colE1 ori only.



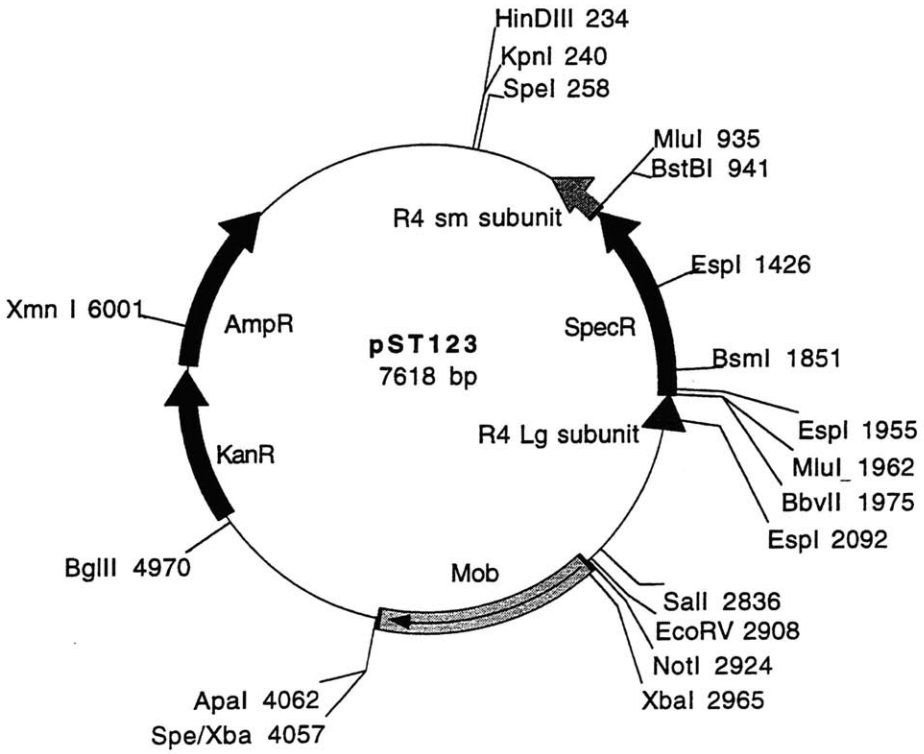
Plasmid name: pST120
Plasmid size: 7005 bp
Constructed by: S.Treadway
Construction date: May 1998
Comments/References: PCR amplified large and small dioxygenase subunits from pR4-10 using AP.21 and AP.23 primers. Cloned 3 kb fragment into pTopo2.1. I24 genomic seq is bewteen EcoRI polylinker sites.



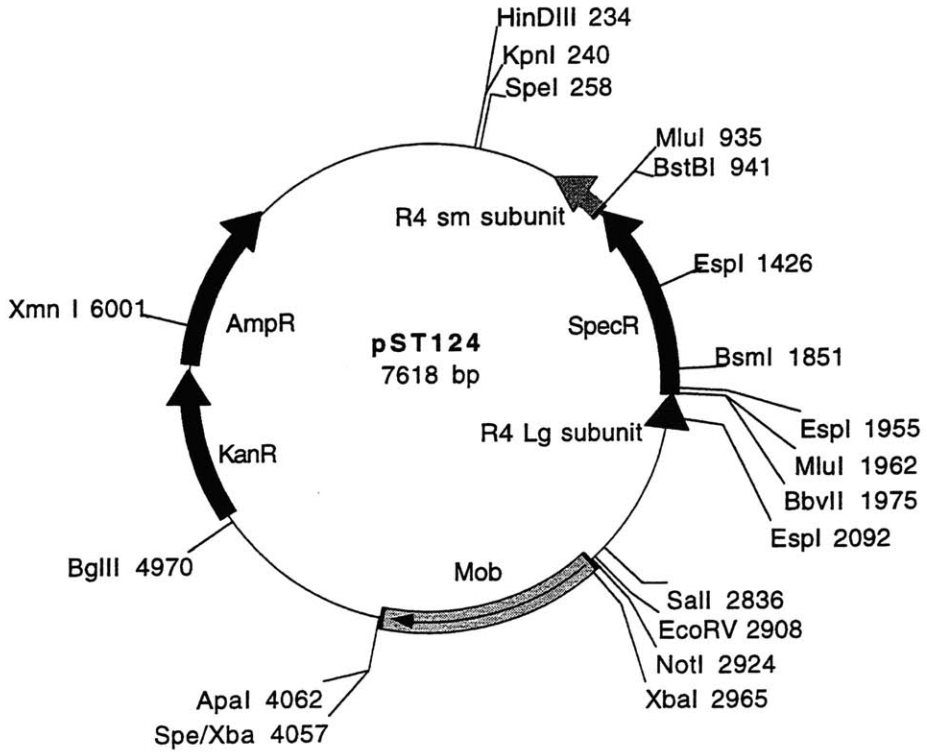
Plasmid name: pST121
Plasmid size: 4947 bp
Constructed by: S. Treaadway
Construction date: June 1998
Comments/References: Used primers NSpec1 and CSpec1 to amplify specR gene from pAPE1b. Cloned into pCR2.1-Topo plasmid from Invitrogen.



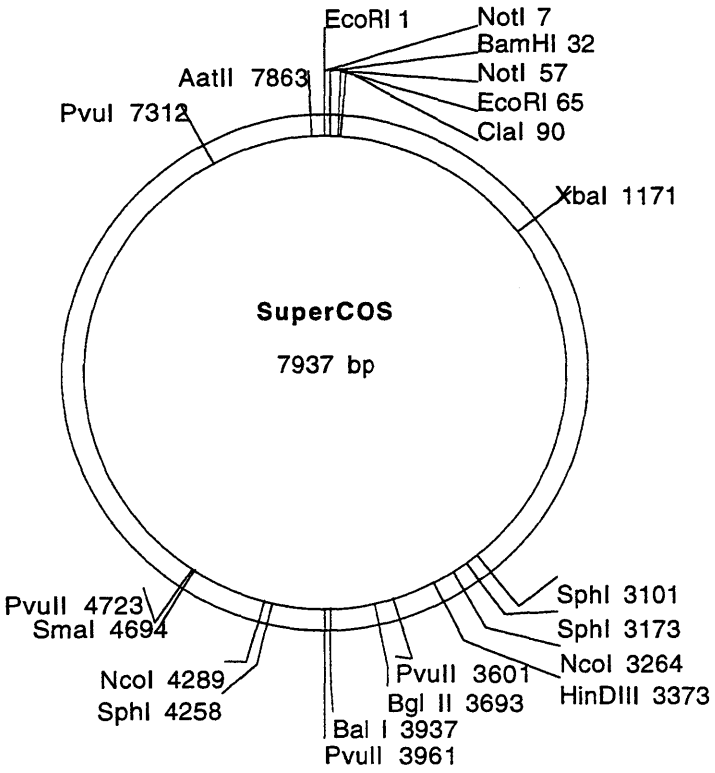
Plasmid name: pST122
Plasmid size: 6505 bp
Constructed by: S. Treadway
Construction date: June 1998
Comments/References: pST121 cut with MluI removing spectinamycin gene. pST120 cut with MluI and SAP treated. Origin: colE1.



Plasmid name: pST123
Plasmid size: 7618 bp
Constructed by: S. Treadway
Construction date: June 1998
Comments/References: pST122 cut with NotI and Xba. Insert is a SpeI-NotI fragment of pAL235 containing the mob element. Contains two spec gene insertions, in tandem.



Plasmid name: pST124
Plasmid size: 7618 bp
Constructed by: S. Treadway
Construction date: June 1998
Comments/References: pST122 cut with NotI and XbaI. Insert is a SpeI-NotI fragment of pAL235 containing the mob element. Knock-out construct of R4 dioxygenase. Origin: colE1.



Plasmid name: SuperCOS
Plasmid size: 7937 bp
Constructed by: Strategene
Construction date: ?
Comments/References: Used to obtain COS sites for new cosmid vector pRhodoCOS.